DOCUMENT RESUME

ED 323 256 TM 015 499

AUTHOR Byrne, Barbara M.; Baron, Pierre

TITLE The Beck Depression Inventory: A Cross-Validated Study of Factorial Structure for Nonclinical Adolescents.

PUB DATE Aug 90

NOTE 48p.; Paper presented at the Annual Meeting of the American Psychological Association (98th, Boston, MA, August 10-14, 1990).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Adolescents; Depression (Psychology); *Factor Structure; Foreign Countries; High Schools; *High School Students; *Predictive Validity; Psychological Testing; *Test Validity

IDENTIFIERS *Beck Depression Inventory; Canada; Confirmatory Factor Analysis; *Cross Validation; Exploratory Factor Analysis

ABSTRACT

Based on three independent samples of high school adolescents in central Canada, aged 12 to 18 years, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to identify, test, and cross-validate first-order and second-order factorial structures underlying the Beck Depression Inventory (BDI). Specifically, the study focused on: (1) determining an initial first-order structure of the BDI based on Group 1 (n=196); (2) postulating and testing alternative first-order structures using CFA procedures based on Group 2 (n=284), on the basis of EFA findings; (3) testing for the validity of a hierarchical structure that includes a single second-order factor of general depression; and (4) cross-validating plausible first-order and second-order structures across Group 3 (n=205) using double cross-validation procedures to identify models having the highest degree of predictive validity. Based on these three increasingly stringent tests of the BDI, the findings suggest that, for non-clinical adolescents, the most appropriate factorial structure comprises a higher-order general factor of depression, with three first-order factors--negative attitudes, performance, and somatic elements. Four data tables are included. (Author/TJH)

The Beck Depression Inventory: A Cross-validated Study of Factorial Structure for Nonclinical Adolescents

Barbara M. Byrne and Pierre Baron
University of Ottawa

Paper presented at the American Psychological Association Annual Meeting, Boston 1990

Abstract

Based on three independent samples of high school adolescents (12-18 yrs; M=15 yrs), exploratory (EFA) and confirmatory (CFA) factor analytic procedures were used to identify, test, and cross-validate 1st- and 2nd-order factorial structures underlying the Beck Depression Inventory (BDI). Specifically, purposes of the study were fourfold: (a) To determine an initial 1st-order structure of the BDI based on Group 1 (N=196), (b) on the basis of EFA findings, to postulate and test alternative 1st-order structures using CFA procedures based on Group 2 (N=284), (c) to test for the validity of an hierarchical structure that includes a single second-order factor of general depression, and (d) to cross-validate plausible 1st- and 2nd-order structures across Group 3 (N=205) using double cross-validation procedures to identify models having the highest degree of predictive validity. Based on these three increasingly stringent tests of the BDI, our findings suggest that, for nonclinical adolescents, the most appropriate factorial structure comprises a higher-order general factor of depression, with three 1st-order factors which we labelled Negative Attitudes, Performance, and Somatic Elements.

The Beck Depression Inventory: A Cross-validated Study of Factorial Structure for Nonclinical Adolescents

The popularity of the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) is unquestioned. Furthermore, despite development of short (Beck & Beck, 1972) and revised (Beck, 1979) versions of the instrument, the original scale remains the most widely used with clinical, as well as nonclinical samples. To date, however, factor analytic research has yielded vastly discrepant findings for both populations. Furthermore, with one exception (Tanaka & Huba, 1984), assessment of factorial validity has been limited to exploratory factor analyses (EFAs), most of which have used principal components procedures and assumed orthogonality among factors; these procedures, however, are widely known to lack the statistical rigor to directly test hypotheses related to factorial structure. The present study addressed these limitations and had four primary purposes: (a) To reexamine the factorial structure of the BDI for nonclinical adolescents using EFA procedures that considered correlated factors, (b) on the basis of EFA findings, to postulate and validate a model of first- order factorial structure with a second independent sample, (c) to test for the validity of an hierarchical structure that includes a single second-order general factor, and (d) to cross-validate plausible first- and second-order models of BDI structure with a third independent sample in order to determine the one yielding the highest level of predictive validity; model fitting, testing, and cross-validation included the Tanaka and Huba (1984, 626) model based on data from Steer and associates (Steer, Shaw, Beck, & Fine, 1977).[1] All analyses based on Samples 2 and 3 used a CFA approach to the data within the framework of covariance structure modeling.

The BDI was originally developed for use with psychiatric populations. Items were derived from clinical observations of symptoms and attitudes frequently displayed by depressed psychiatric patients, and sometimes exhibited by those considered to be nondepressed (Beck, Steers, & Garbin, 1988). The psychometric properties of the BDI, as they bear on various clinical populations, have undergone extensive testing (for a review, see Beck et al., 1988). Overall, results have been consistent in reporting high reliability and moderately strong validity. Tests of internal consistency reliability have yielded alpha coefficients ranging from .79 (Strober, Green, & Carlson, 1981) to .95 (Beck et al., 1988) (M=.87), a split-half coefficient of .93 (Beck et al., 1988), and a KR-20 coefficient of .53 (Weckowicz, Muir, & Cropley, 1967); test-retest reliability coefficients have ranged from .48 to .82 (Beck et al., 1988) (M=.66), with time lags extending from 5 days to 1 month. Reported convergent validity coefficients have ranged from .55 to .73 (M=.64) (Beck et al., 1988) with clinical ratings (M=.64), and from .33 (Beck, Carlson, Russell, & Brownfield, 1987) to .86 (Beck et al., 1988) (M=.49) with other scale rating measures of depression.

In contrast, findings from research bearing on the factorial structure of the BDI, when used with clinical populations, has been markedly less consistent. Indeed, researchers have reported optimal solutions ranging from one through seven factors. Except for the Tanaka and Huba (1984), all studies have employed an exploratory mode of inquiry using either Common Factor (Cropley & Weckowicz, 1966; Steer et al., 1977; Steer, McElroy, & Beck, 1982), or principal components (Beck & Lester, 1973; Gibson & Becker, 1973; Lester & Beck, 1977; Reynold & Gould, 1981; Shaw, Steer, Beck, & Schut, 1979; Vrendenburg & Flett, 1985; Weckowicz et al., 1967) analyses; all have employed varimax rotation procedures thereby imposing the restriction of factor independence.

Of particular importance to the present study, however, is the work of Tanaka and Huba (1984) who used CFA procedures to test for a higher-order factorial structure. Based on two separate clinical samples derived from the earlier data of Beck and Lester (1973) and Steer et al. (1977), Tanaka and Huba concluded the latent structure of the BDI to be hierarchical; as such, they argued for a single second-order factor of General Depression at the apex which, in turn impacted on three, more specific first-order factors which they labelled Negative Attitudes/Suicide, Performance Difficulty, and

Physiological. In the present study, we test and cross-validate their model of the Steer et al. data for use with normal adolescents.

Recently, there has been increasing interest in using the BDI with nonclinical populations. Relatedly, in a search for evidence in support of such applications, a number of researchers have tested both its psychometric properties and its factorial structure with a variety of nonclinical samples; these have included college students (Golin & Hartz, 1979; Gould, 1982; Lips & Ng, 1985; Tanaka-Matsumi & Kameoka, 1986), high school students (Barrera & Garrison-Jones, 1988; Baron & Laplante, 1984; Hill & Kemp-Wheeler, 1986; Kauth & Zettle, 1990; Teri, 1982), male cardiac patients (Campbell, Burgess, & Finch, 1984), expectant parents (Lips & Ng, 1985), and adults in their middle (Lips & Ng, 1985) and senior (Gallagher, Nies, & Thompson, 1982); additionally, a Chinese edition of the BDI has been validated for 2,150 secondary school students (Shek, 1990).

Internal consistency reliability of the BDI for these nonclinical populations has revealed alpha coefficients ranging from .73 (Gallagher et al., 1982) to .90 (Barrera & Garrison-Jones, 1988) (M=.83), Spearman-Brown coefficients from .53 (Gallagher et al., 1982) to .95 (Beck et al., 1988) (M=.78), and a KR-20 coefficient of .52 (Golin & Hartz, 1979); test-retest relibilities have ranged from .62 to .83 (M=.73) over periods of 6 hours to four months (Beck et al., 1988).

Reported validity coefficients have ranged from .33 (Beck et al., 1988) to .91 (Tanaka-Matsumi & Kameoka, 1986 (M=.73) for convergence with other scale measures of depression, and a coefficient of .77 with physicians' ratings (Bumberry, Oliver, & McClure, 1978).

While these findings have been fairly consistent in substantiating the psychometric adequacy of the BDI for use with nonclinical populations, those from research bearing on its factorial structure, as was the case with clinical populations, have been extremely discordant. Again, results have argued for underlying structures composed of a single, through a total of seven factors, with method of extraction limited largely to principal components analyses using varimax rotation.

Taken together, findings related to reliability and concurrent validity provide strong evidence that the BDI is psychometrically sound in its measurement of depression for both clinical and nonclinical populations. Construct validity research bearing on its factorial structure, however, is much less glowing. Thus, the generalizability of these studies is severely limited for several reasons. First, except for one study (Tanaka & Huba, 1984), factorial validity of the BDI has been examined using only an exploratory approach based, for the most part, on principal components analyses. Now well documented, however, are several deficiencies associated with such exploratory procedures in general (see e.g., Bollen, 1989;

Long, 1983; Marsh & Hocevar, 1985), and principal components analysis in particular (see e.g., Borgatta, Kercher, & Stull, 1986; Gorsuch, 1990; Hubbard & Allen, 1987; Snook & Gorsuch, 1989). Indeed, the latter has been shown to yield highly inflated factor loadings and, thus, misleadingly clear factor structures. CFA, on the other hand, can provide more powerful tests of factorial validity.

Second, virtually all of the EFA work, to date, has assumed the independence of multiple factors in explaining the covariation of symptoms. Tanaka and Huba (1984), however, demonstrated that once the restriction of factor orthogonality was removed, the primary level factors of the BDI were highly intercorrelated; such findings suggest the presence of a higher-order general factor.

Third, although Tanaka and Huba (1984) cross-validated their findings of a higher-order factor structure with a second independent clinical sample (Steer et al., 1977), they were prevented from using CFA procedures to test for the equivalency of factor structure across samples due to the availability of data in correlation, rather than in covariance matrix format. Although they obtained a congruence of similarity coefficient of .961 between the solutions in the two samples, this procedure provides a much weaker test of factorial invariance. A substantially more rigorous approach would involve cross-validation with an independent sample using a simultaneous analysis of data.

Finally, construct validity research bearing on the BDI, for use with nonclinical populations, is completely void of studies that have taken an a priori approach to testing for factorial structure. In light of rapidly increasing use of this assessment measure with student populations at both the secondary and postseconday levels, there is clearly a need to validate the instrument using statistically appropriate and rigorous testing procedures.

The present study addressed these concerns by (a) testing the validity of the BDI for secondary school adolescents using both EFA and CFA approaches to the validation inquiry, (b) testing for the validity of an hierarchical model of factorial structure, (c) conducting double cross-validation procedures to determine the factorial model having the greatest predictive power, and (d) using one of three independent samples of adolescents from the same population in the application of each of the foregoing procedures; in cross-validation terms, Samples 1 and 2 operated as calibration samples; Sample 3 provided the validation sample.

Method

Sample and Procedures

The data comprised BDI responses from three independent samples of adolescents in grades 9 to 12, all of whom attended the same secondary school in a large metropolitan area in central Canada. Participants ranged in age from 12 to 18 years and may be described as follows: Group 1 (n=196; 109 males, 87 females; mean age = 14 yrs), Group 2 ($\underline{n}$=284; 137 males, 147 females; mean age = 15 yrs), Group 3 ($\underline{n}$=205; 105 males, 100 females; mean age = 15 yrs); all groups were drawn from earlier studies designed to identify various correlates symptomatic of depression in high school adolescents (see Baron, 1988, 1989; Baron & MacGillivray, 1989). Listwise deletion of missing data resulted in final sample sizes of 179, 284, and 189 for Groups 1, 2, and 3, respectively.

In each instance, subjects completed the BDI, along with other assessment measures, during one regular class period; all testing materials were completed anonymously. Test instructions were paraphrased by the test administrator, and procedural questions were solicited and answered. All participation, in keeping with school and Ethics Committee policies, was voluntary, and no incentives were offered; subjects were provided with a brief rationale for the study.

Instrumentation

The BDI is a 21-item scale that measures symptoms and attitudes related to cognitive, behavioral, affective, and somatic components of depression. Although originally designed for use by trained interviewers, it is now most typically used as a self-report measure (Beck et al., 1988; Kearns, Cruickshaw, McGuigan, Riley, Shaw, & Snaith, 1982; Vredenburg et al., 1985). For each item, respondents are presented with four statements rated from 0 to 3 in terms of intensity, and asked to select the one which most accurately describes their own feelings; high scores representing a more reported depression. Total scores range from 0 to 63 and are used to categorize four levels of depression: none to minimal (0-9), mild to moderate (10-18), moderate to severe (19-29), and severe (30-63) (Beck et al., 1988).

Data Analyses

Data were analyzed in four stages. First, EFAs of the data for Group 1 were conducted using Maximum Likelihood extraction with oblique rotation. Second, based on the optimal simple structure determined in Stage 1 of the analyses a first-order factorial model of the BDI was specified a priori, and its validity tested statistically on data from Group 2 using analysis of covariance structures within the framework of a CFA model; these analyses were based on the LISREL VI computer program (Joreskog & Sorbom, 1985). Third, given evidence of substantial factor correlations, as determined by the exploratory analyses in Stage 1, an hierarchical factor analysis of the data from Group 2 was conducted to test for the presence of a higher-order factorial structure; the second-order model proposed by Tanaka and Huba (1984) was also put to the test statistically. Finally, statistically best-fitting, and theoretically plausible first- and second-order models were tested across calibration (Group 2) and validation (Group 3) samples using procedures of double cross-validation to determine the model yielding the highest degree of predictive validity (see Cudeck & Browne, 1983; Bagozzi & Yi, 1988).

Although, historically, the chi square (χ^2) likelihood ratio has been used to judge the extent to which a proposed covariance structure model fits the observed data, the sensitivity of the χ^2 statistic to sample size is now well documented; several alternative nonstatistical (i.e., practical) indices of fit have subsequently been proposed (for a recent review, see Marsh, Balla, & McDonald, 1988). As recommended by a number of researchers (e.g., Bagozzi & Yi, 1988; Bollen, 1989; Joreskog, 1982; Long, 1983; Marsh & Hocevar, 1985), assessment of fit in the present study was based on multiple criteria that reflect statistical, theoretical, and practical considerations. Furthermore, parsimony, as well as degree of fit were taken into account (see Bentler & Mooijaart, 1989; Mulaik, James, Alstine, Bennett, Lind, & Stilwell, 1989). As such, evaluation of model fit was based on (a) the χ^2 likelihood ratio, (b) the Bentler (1990) revised normed comparative fit index (CFI), (c) the parsimonious normed comparative fit index (PCFI; Mulaik et al., 1989), (d) the adjusted goodeness-of-fit index (AGFI), T-values (parameter estimates relative to their standard errors of estimate), and modification indices, all provided by the LISREL program, and (e) the substantive meaningfulness of the model (see MacCallum, 1986; Suyapa, Silvia, & MacCallum, 1988).

The CFI represents a modified version of the Bentler and

Bonett (1980) normed fit index that has been corrected for sample size dependency (Bentler, 1990); it is included here because of widespread use of the former index. Plausible models yield CFI values >.90, indicating a reasonably adequate approximation to the observed data. The PCFI is based on the CFI; the originally proposed index (Mulaik et al., 1989) has been adapted here for use with the CFI. Because this index takes into account goodness-of-fit, as well as the parsimony of the model, its value is often much lower than what is generally perceived as "acceptable" on the basis of normed indices of fit. In fact, Mulaik and associates (1989) note that goodness-of-fit indices in the .90's accompanied by parsimonious fit indices in the .50's are not unexpected. For purposes of the present study, however, PCFI values >.80 were considered indicative of an adequately specified model (see e.g., Mulaik et al., 1989).

Results and Discussion

Exploratory Factor Analyses: Group 1

Taking into account the range of EFA structures reported in the literature, common factor analyses were conducted for potential 2-, through 7-factor solutions; considering sample size and case/variable ratio, a value of .35 was used as the cutpoint for judging the saliency of factor loadings (see Gorsuch, 1983). The 2-factor solution, as well those based on more than four factors were rejected for several reasons: (a) the factor patterns were notably ill-defined and could not be interpreted meaningfully, (b) the pattern of loadings changed drastically from one solution to the other, and (c) goodness-of-fit for each model, as indicated by the χ^2 statistic as well as other statistical criteria, was less than adequate. Consequently, findings from only the 3- and 4-factor solutions are now described, the less plausible of the two models being presented first.

Four-factor Solution

Interestingly, within the framework of the 4-factor solution, a pattern of loadings emerged that bore close resemblance to those modeled by Tanaka and Huba (1984). As such, eight items with loadings ranging from .27 to .76 ($M=.51$) were consistent with the factor the authors labelled Negative Attitudes/Suicide. Additionally, three items were consistent with their Physiological factor, and two with their factor labelled Performance Difficulty; the fourth factor shared salient loadings representative of Tanaka and Huba's Negative Attitudes/Suicide and Performance Difficulty factors. Once again, this model was rejected on the basis of substantive, as well as statistical criteria.

Although there was a hint of potentially meaningful structure in this solution, the pattern was marred by innumerable cross-loadings (>.20) which obviated any clear factorial structure that could lend itself to meaningful interpretation. Indeed, the pattern of fragmentation was clearly indicative of an overfactored solution (see Gorsuch, 1983; Walkey, 1983).

In terms of goodness-of-fit, however, the 4-factor solution was statistically better fitting (χ^2 (132) = 139.45) than the 3-factor solution (χ^2 (150) = 174.58), albeit both represented statistically plausible models (p >.05). Nonetheless, two additional statistical criteria argued against the 4-factor solution. First, examination of the variance statistics revealed that the fourth factor contributed only 2.5% to the total amount extracted (31.5%). Second, the scree plot of eigenvalues (Cattell, 1966) did not clearly demarcate four factors. For these statistical reasons, and because substantive clarity of factors is critical to meaningful interpretation of item responses from assessment scales (see Walkey, 1983), the 4-factor solution was rejected in favor of a 3-factor structure; we turn now to these results.

Three-factor Solution

In contrast to the previous solution, factor pattern for the 3-factor solution was remarkably similar to the Tanaka and Huba (1984) model; 15 of 21 loadings replicated the Steer et al. data, and 18 of 21 the Beck & Lester data. Specifically, nine items loaded on the Tanaka and Huba Negative Attitudes/ Suicide factor having loadings that ranged from .24 to .85, (M=.43), seven items loaded on the Performance Difficulty factor (loadings from .24 to .65 (M=.52), and two loaded on the Physiological factor. Results for the 3-factor solution, in total, are reported in Table 1.

Insert Table 1 about here

Although factor loadings for the Physiological factor might appear at first blush to be somewhat aberrant, they are not totally unexpected based on reported findings from other factor analytic studies of the BDI (e.g., Teri, 1982); the items (#18, #19) represent appetite loss and weight loss, respectively. The fact that Item #18 loaded >1.00 and Item #19 <.18 is a statistical artifact arising from multicollinearity, and is a clear indication of item redundancy; both items are eliciting essentially the same information from respondents. Of note is the suggestion by some authors (e.g., Kauth & Zettle, 1990) that Item # 21 (libido loss) should be eliminated when using the BDI with adolescents; loading for this item with the present samples (as will be noted later), was not unreasonable. Finally, as expected, the three factors were substantially related, with factor correlations ranging from .25 to .55 (M=.39).

Although the EFA findings suggest that a 3-dimensional stucture underlies the BDI for nonclinical adolescents, the true test must come from a CFA approach to the data, since EFA is limited in its ability to: (a) yield unique factorial solutions, (b) define a testable model, (c) assess the extent to which an hypothesized model fits the observed data and suggest alternative parameterization for model improvement, and

(d) adequately test for factor equivalence across groups (Bollen, 1989; Fornell, 1983; Long, 1983; Marsh & Hocevar, 1985). In contrast, CFA can yield this information and is therefore a more powerful test of factorial validity. We turn now to the CFA results as they bear on our second sample of adolescents.

Confirmatory Factor Analyses: Group 2

Since our intent in this study was to cross-validate findings across a third independent sample, we consider Group 2 as our calibration sample and take a post hoc CFA approach to the analyses. That is, for both our 1st- and 2nd-order factor analyses, we first specify a restricted model of BDI structure consistent with our EFA findings, and then seek to improve the fit of this model through respecification based on related modification indices (MIs)[2] provided by the LISREL program. (For a more extensive discussion and application of these procedures, see Byrne, 1989.) As such, we maintain our exploratory mode of inquiry while concomitantly benefiting from both a more stringent testing of an hypothesized 3-factor structure, and the provision of additional information made possible by the CFA data analytic approach (see Anderson & Gerbing, 1988; MacCallum, 1986). We turn first to our 1st-order factor analyses.

First-order Factorial Structures

In keeping with our 3-factor EFA findings, the initial CFA model hypothesized a priori that: (a) responses to the BDI could be explained by three factors of depression which we labelled Negative Attitudes, Performance Difficulty, and Somatic Elements, (b) each item would have a non-zero loading on the depression factor it was hypothesized to measure, and zero loadings on all other factors, (c) the three factors would be correlated and, (d) the uniqueness terms for the item variables would be uncorrelated. (The term "uniqueness" is used here in the factor analytic sense to mean a composite of specific and random measurement error which, in cross-sectional studies, cannot be separated; see Gerbing & Anderson, 1984). Because our study involved multigroup comparisons (see Byrne, Shavelson, & Muthén, 1989; Marsh & Hocevar, 1985) and because of possible distortion to parameter estimates (Bollen, 1989; Boomsma, 1983), analyses were based on covariance, rather than on correlation matrices.[3] Results for these CFA analyses are presented in Table 2.

Insert Table 2 about here

As indicated in Table 2, goodness-of-fit for our initially hypothesized 3-factor structure (Model 1) was less than adequate from both a statistical (χ^2 value), and a practical (AGFI, CFI, PCFI values) perspective; this model was therefore rejected.[4] (Model 0 argues that each item represents a factor and represents the null model needed for computation of the CFI.) Thus, we proceeded next to determine the 3-factor model that most plausibly represented the observed data for Group 2. Inspection of the MIs indicated that by allowing Item #8 to cross-load onto Factor 1 (Negative Attitudes), as well as on Factor 2 (Performance Difficulty), would lead to a substantial improvement in model fit. Model 2 was therefore respecified and reestimated accordingly.

To compare the difference between Models 1 and 2, we evaluate the difference in χ^2 ($\Delta\chi^2$) between the two models. Since this differential is itself χ^2-distributed, with degrees of freedom equal to the difference in degrees of freedom (Δdf), it can be tested statistically; a significant $\Delta\chi^2$ indicates a substantial improvement in model fit. Reviewing the results in Table 2, we see evidence of a highly significant improvement in fit between Models 1 and 2. Nonetheless, examination of the MIs pointed to the possibility of still further improvement in fit through the relaxation of particular parameters. Thus, model-fitting proceeded until we were satisfied that all substantively plausible models had been identified.[5] The remaining content of Table 2, then, summarizes these analyses.

As indicated in Table 2, all subsequently specified models demonstrated a significant improvement in fit over the initially hypothesized model. While the results of these analyses are fairly straightforward, a few words of explanation regarding the rationale underlying the specification of two models (Model 8, Model 10) are, perhaps, in order. Accordingly, given the desirability of parsimony in model fitting, it was deemed important to determine whether or not the cross-loadings of Items #9, #16, and #21 onto Factor 3 were statistically sound. Thus, a model in which these parameters were deleted was specified (Model 8). As shown in Table 2, this model resulted in a highly significant decriment to model fit. Given the known importance of these parameters to the model, it was then important to test the validity of their sole loading onto Factor 3; this model Model 9) was then tested and found to yield a slightly greater decrement in model fit, albeit the fit indices remained unchanged. Finally, one consequence of the respecification of Items #8 and #12 on Factor 1 was a substantial increase in the MI value for Item #9 as it related to Factor 1. Thus, an additional model was estimated (Model 10) in which this item was specified as loading solely onto Factor 1; the fit of this model demonstrated a fairly large improvement over that of Model 9.

Overall, then, given considerations of statistical and subjective fit, parsimony, and substantiveness meaningfulness, certain models became clearly more appropriate than others with respect to our adolescent population. For example, although Model 6 represented the best statistical fit to the data, it was the least parsimonious, which makes it less desirable than say, Model 7, albeit given the same degree of subjective goodness-of-fit (CI=.93); this value indicating that 93% of the covariation in the data is explained by this particular model. Mindful of these important tradeoff factors, as well as the known sensitivity of the X^2 statistic to sample size, we favor Model 10 as the most appropriate model; Model 9, nonetheless, is certainly a viable alternative. The true test of model appropiateness, however, must come from its power of prediction in a second independent sample; thus, final decision regarding the most desirable 1st-order model rests on results from the cross-validation analyses (to be described later). We turn now to results related to our tests of 2nd-order models.

Second-order Factorial Structures

In order to test the hypothesis of a higher-order factorial structure underlying the BDI, as suggested by Tanaka and Huba (1984), we respecified our EFA model as a 2nd-order factor model. As such, we postulated a single higher-order factor representing general depression, and three lower-order factors representing as before, Negative Attitudes, Performance Difficulty, and Somatic Elements. As with the 1st-order factor analyses, we first tested a restricted model of our original EFA solution and then proceeded in a post hoc manner to respecify and reestimate alternative models in search for a well-fitting factorial structure. Additionally, we tested the validity of the Tanaka and Huba model for use with adolescents. Results for these analyses are presented in Table 3.

Insert Table 3 about here

Interestingly, all but one of the alternative models remained consistent with those specified within the framework of a 1st-order structure. In this regard, contrary to findings from the 1st-order factor analyses, Item #12 measuring Withdrawal, showed no indication of either cross-loading or fully loading onto Factor 1; rather, it loaded significantly on the Performance Difficulty factor, thus differentiating it also, from the Tanaka and Huba (1984) model. Based on the same rationale cited earlier with regard to 1st-order structures, we again favor our final model (Model 9 in Table 3) as the most plausible model of factorial structure for the the BDI when used with a nonclinical adolescent population; again, Model 8 (Model 9 in Table 2) must be considered a worthy competitor.

In sum, we determined several plausible factorial models that potentially underlie the BDI when used with a nonclinical adolescent population. The task now was to ascertain which of these models best replicated across our third group of adolescents. We turn now to these analyses.

Cross-validation Analyses: Group 3

Support for a factorial model comes in its ability to replicate across an independent sample. Having identified four plausible factorial models within the frameworks of both 1st-order and 2nd-order structures, our task now was to test for their the predictive accuracy across a second independent sample of adolescents. As noted by Bagozzi and Yi (1988), cross-validation is best regarded as a method for comparing alternative models, with the selection of those that most accurately approximate the underlying population structure.

Of course, the most stringent assessment of competing models is accomplished by means of double cross-validation procedures (see Bagozzi & Yi, 1988; Cudeck & Browne, 1983); this was the approach used in the present study. Only the most plausible models were tested in these cross-validation analyses; plausibility was based on parsimony, as well as good-of-fit to the data. As such, four 1st-order Models (7, 8, 9, 10; see Table 2), and four 2nd-order Models (6, 7, 8, and 9; see Table 3) were estimated for the validation sample (Group 3), as well as for the calibration sample (Group 2); for sake of completeness, the original EFA-specified model for both 1st- and 2nd-order structures, and the Tanaka and Huba (1984) 2nd-order model were both tested. Analyses proceeded in a 2-step manner. First, the parameters derived from the calibration sample were used for prediction in the validation sample, and the goodness-of-fit criteria derived. Second, these analyses were repeated, but with the rc es of the calibration and validation samples reversed. For example, 1st-order Model 7 was estimated for the validation sample (Group 3) with the item measurement and factor covariance parameter estimates from the calibration sample (Group 2) fixed a priori; Model 7 was then estimated for the calibration sample with these same parameters fixed a priori to the values estimated for the validation sample.

With the introduction of their cross-validation index

(CVI), Cudeck and Browne (1983) provided a means to evaluating cross-validation analyses within the framework of covariance structure models with latent variables. The CVI represents the discrepancy between the sample covariance matrix of one sample (say, the calibration sample), and the restricted covariance matrix as imposed on a second independent sample (say, the validation sample). The model that exhibits the smallest CVI in each of the two sets of cross-validation analyses, is considered the one with the highest degree of predictive validity. Results based on these procedures are presented in Table 4.

Insert Table 4 about here

Turning first to results related to 1st order structures, we see that Model 9 demonstrated the greatest predictive validity across independent samples. Although this model did not represent the best-fitting model, it was more parsimonious which likely led to its higher degree of predictive power.[6] Within the context of a 1st-order factorial structure, these findings argue for a 3-factor structure underlying the BDI, with a pattern of factor loadings as presented in Figure 1.

Insert Figure 1 about here

As an aid to interpretation, a few words of explanation are in order. Factors are enclosed in circles, and item measurements enclosed in rectangles; the single-headed arrows linking each of the circles to a set of rectangles represent the regression of the underlying latent construct on a particular item. The double-headed arrows represent correlations among the factors. Finally, the single-headed arrows leading to each of the circles represent residual error in the equation; those leading to the rectangles represent error/uniqueness associated with the observed measurements. Except for weight loss,[7] all factor loadings and intercorrelations were statistically significant, as indicated by the parenthesized critical ratio values; these values are distributed as a z-statistic and, thus, those >1.96 are indicative of statistical significance.

Aside from the fact that the model in Figure 1 represents a 1st-order structure, its basic pattern of loadings is markedly similar to the one presented by Tanaka and Huba (1984) for the Steer et al. (1977) data. In particular, six items loaded differently for our adolescent sample: self-image, dissatisfaction, irritability, suicidal ideation, insomnia, and libido loss. That for these adolescents, self-image and libido loss loaded onto the Negative Attitudes and Somatic Elements factors, respectively, seems subtantively reasonable; given the clinical sample of alcoholics comprising the Steer et al data, such loading differences are not unexpected. Of additional import is the strength of association among each of the latent factors;[8] these correlations, as emphasized by Tanaka and Huba (1984), argue for a hierarchical factorial structure.

Turning now to results as they bear on a 2nd-order structure, we see that the model exhibiting the lowest CVI, and therefore demonstrating the highest degree of predictive validity, is Model 9 (Model 10 for 1st-order models). Again, although this model was not the best-fitting one, it was more parsimonious than Model 6, in which there were three cross-loaded items; additionally, and possibly most important, it was considered more theoretically sound. Model 9 is presented schematically in Figure 2.

Insert Figure 2 about here

In sum, the reformulation of our 1st-order factor models into 2nd-order structures yielded two important differences. First, whereas Item #12 measuring Withdrawal loaded onto Factor 1 (Negative Attitudes) when specified within the framework of a 1st-order factor model, this item loaded onto Factor 2 (Performance Difficulty) when respecified as a 2nd-order factor model. Second, although Item #9, measuring suicidal ideation, demonstrated a tendency to load on Factor 1 instead of Factor 3 in both sets of analyses, this item when specified within the framework of a 2nd-order structure, proved to be the one that cross-validated best across the third group of adolescents; in constrast, Item #9 loaded on the Somatic Elements factor when specified as a 1st-order factorial structure. Specification as a 2nd-order structure, then. made the loading of Item #9 consistent with the Tanaka and Huba model.

Finally, results from tests of the Tanaka and Huba (1984) model of the Steer et al. (1977) data demonstrated a weaker ability to cross-validate across an independent sample, thus casting doubt on the replicability of their proposed model of BDI factorial structure, at least when used with nonclinical adolescents. Given that the Steer et al. data were based on item responses from a clinical sample of alcoholic black adults, these findings are not unexpected. Indeed, the fact that the pattern of loadings remained very similar to the Tanaka and Huba model must be considered, in itself, quite remarkable!

Conclusion

In broad terms. our ultimate aim in the present study was to determine a rigorously tested and cross-validated factorial structure for the BDI, with generalizability to the nonclinical adolescent population. More specifically, we tested the factorial structure of depression underlying the BDI by incorporating the use of both exploratory and confirmatory factor analytic procedures using three independent samples. EFA was used with the first sample to generate an initial factorial model; CFA was used with the remaining two samples. Both 1st- and 2nd-order factorial structures were specified for the second sample, and then tested, statistically, for their goodness-of-fit. Finally, the most plausible models, within the framework of both 1st- and 2nd-order structures, were cross-validated across a third independent sample. These analyses, we believe, provide the most statistically stringent test of the BDI to date, with respect to its use with normal adolescents.

On the strength of our results, it seems evident that the BDI, when used with normal adolescents, is most adequately defined by a 3-factor, hierarchical structure of depression. From a statistical perspective there was little to choose between the 1st- and 2nd-order structures. Considered within the framework of theory, however, the higher-order structure represented a more conceptually rational explanation of the data. Indeed, the alternative loadings of withdrawal and suicidal ideation are particularly noteworthy. Clearly, the symptom of withdrawal is indicative of social inhibition as a consequence of experienced difficulty in responding to the demands of social activities; its loading on the Performance Difficulty within a 2nd-order structure, therefore, is considered reasonable. The loading of withdrawal on the Negative Attitudes factor, on the other hand, can perhaps be explained by the globality of the 3-factor solution which may to some extent hamper a better delineation of the diverse attitudinal symptoms of depression.

The loading of suicidal ideation on the Negative Attitudes factor in the 2nd-order structure is decidedly more reasonable than its alternative loading onto the Somatic Elements factor, as was the case in the 1st-order structure. We are at a loss to explain the latter result. but consider it to be an artifact of sampling variability; thus, we expect this finding to disappear upon further replication.

Although the aberrant loading values of weight loss are indicative of multicollinearity with appetite loss, additional explanation possibly lies with the scoring of this item, and with its perceived interpretation by normal adolescents. Scoring may be problematic in that given a 'yes' response to the statement "I am purposely trying to lose weight by eating less", the item was automatically scored as zero, indicating no depression. Beyond this, however, one must consider the importance of body image during the adolescent years. In this regard, the enormous socialization impact of the media is unquestioned. Consequently, admission of weight loss in response to this item may, indeed, be the result of many factors other than depression.

In light of the rising incidence of broken homes, physical and sexual abuse, drug and alcohol abuse, and other pathological ills present in today's society, there is growing concern among school psychologists, counselors, and administrators that many adolescents may be at severe risk for depression. Given the increasing popularity of the BDI for use with nonclinical populations, and because meaningful interpretations of test scores are always contingent on the factorial soundness of the measuring instrument, we considered it important to validate its use with normal high school adolescents. We expect our findings to be of substantial interest to academicians, as well as to various professionals concerned with the well-being of adolescents. Academically, findings should provide important links to future research that focuses on methodological and substantive issues related to the BDI as it bears on the nonclinical adolescent population. From a practical perspective, results of the study should carry important implications for the application of the BDI, and for the interpretation of its test scores relative to this population.

References

Anderson, J.C., & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103, 411-423.

Bagozzi, R.P., & Yi, Y. (1988). On the evaluation of structural equation models. Academy of Marketing Science, 16, 74-94.

Baron, P. (1986). Egocentrism and depressive symptomatology in adolescents. Journal of Adolescent Research, 1, 431-437.

Baron, P. (1988). Rôle sexuel et symptômes dépressifs chez les adolescents. Comportement Humain, 2, 81-85.

Baron, P., & LaPlante, L. (1984). L'inventaire de Dépression de Beck: Son utilisation auprès d'un échantillon d'adolescents francophones. Revue de Modification du Comportement, 14, 161-166.

Baron, P., & MacGillivray, R.G. (1989). Depressive symptoms in adolescents as a function of perceived parental behavior. Journal of Adolescent Research, 4, 50-62.

Barrera, M. Jr., & Garrison-Jones, C.V. (1988). Properties of the Beck Depression Inventory as a screening measure for adolescent depression. Journal of Abnormal Child Psychology, 16, 263-273.

Beck, A.T., & Beck, R.W. (1972). Screening depressed patients in family practice: A rapid technique. Postgraduate Medicine, 52, 81-85.

Beck, D.C., Carlson, G.A., Russell, A.T., & Brownfield, F.E. (1987). Use of depression rating instruments in developmentally and educationally delayed adolescents. Journal of American Academy of Child and Adolescent Psychiatry, 26, 97-100.

Beck, A.T., & Lester, D. (1973). Components of depression in attempted suicides. Journal of Psychology, 85, 257-260.

Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979). Cognitive theory of depression. New York: Guildford Press.

Beck, A.T., Steer, R.A., & Garbin, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. Clinical Psychology Review, 8, 77-100.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-571.

Bentler, P.M. (1990). Comparative fit indices in structural models. Psychological Bulletin, 107, 238-246.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bentler, P.M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. Psychological Bulletin, 106, 315-317.

Bollen, K.A. (1989). Structural equations with latent variables. New York: Wiley.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. Psychometrika, 50, 229-242.

Borgatta, E.F., Kercher, K., & Stull, D.E. (1986). A cautionary note on the use of principal components analysis. Sociological Methods & Research, 15, 160-168.

Bumberry, W., Oliver, J.M., & McClure, J.W. (1978). Validation of the Beck Depression Inventory in a university population using psychiatric estimates as the criterion. Journal of Consulting and Clinical Psychology, 46, 150-155.

Byrne, B.M. (1989). A primer for LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer-Verlag.

Byrne, B.M., Shavelson, R.J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105, 456-466.

Campbell, I.M., Burgess, P.M., & Finch, S.J. (1984). A factorial analysis of BDI scores. Journal of Clinical Psychology, 40, 992-996.

Cattell, R.B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.

Cropley, A.I., & Weckowicz, T.E. (1966). The dimensionality of clinical depression. Australian Journal of Psychology, 18, 18-25.

Cudeck, R., & Browne, M.W. (1983). Cross-validation of covariance structures. Multivariate Behavioral Research, 18, 147-167.

Fornell, C. (1983). Issues in the application of covariance structure analysis: A comment. Journal of Consumer Research, 9, 443-448.

Gallagher, D., Nies, G., & Thompson, L.W. (1982). Reliability of the Beck Depression Inventory with older adults. Journal of Consulting and Clinical Psychology, 50, 152-153.

Gerbing, D.W., & Anderson, J.C. (1984). On the meaning of within-factor correlated measurement errors. Journal of Consumer Research, 11, 572-580.

Gibson, S., & Becker, J. (1973). Alcoholism and depression: The factor structure of alcoholics' responses to depression inventories. Quarterly Journal of Studies in Alcoholism, 34, 400-408.

Golin, S., & Hartz, M.A. (1979). A factor analysis of the Beck Depression Inventory in a mildly depressed population. Journal of Clinical Psychology, 35, 322-325.

Gorsuch, R.L. (1983). Factor analysis. Hillsdale, N.J.: Erlbaum.

Gorsuch, R.L. (1990). Common factor analysis versus component analysis: Some well and little known facts. Multivariate Behavioral Research, 25, 33-39.

Gould, J. (1982). A psychometric investigation of the standard and short-form Beck Depression Inventory. Psychological Reports, 51, 1167-1170.

Hill, A.B., & Kemp-Wheeler, S.M. (1986). What does the Beck Depression Inventory measure in students? Personality and Individual Differences, 7, 39-47.

Hubbard, R., & Allen, S.J. (1987). A cautionary note on the use of principal components analysis: Supportive empirical evidence. Sociological Methods & Research, 16, 301-308.

Joreskog, K.G. (1982). Analysis of covariance structures. In C. Fornell (Ed.), A second generation of multivariate analysis Vol 1: Methods (pp. 200-242). New York: Praeger.

Joreskog, K.G., & Sorböm, D. (1985). LISREL VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Mooresville, IN: Scientific Software.

Kauth, M.R., & Zettle, R.D. (1990). Validation of depression measures in adolescent populations. Journal of Clinical Psychology, 46, 291-295.

Kearns, N.P., Cruickshank, C.A., McGuigan, K.J., Riley, S.A., Shaw, S.P., & Snaith, R.P. (1982). A comparison of depression rating scales. British Journal of Psychiatry, 141, 45-49.

Lester, D., & Beck, A.T. (1977). Suicidal wishes and depression in suicidal ideators: A comparison with attempted suicides. Journal of Clinical Psychology, 33, 92-94.

Lips, H.M., & Ng, M. (1985). Use of the Beck Depression Inventory with three non-clinical populations. Canadian Journal of Behavioural Science, 18, 62-74.

Long, J.S. (1983). Confirmatory factor analysis. Beverly Hills, CA: Sage.

MacCallum, R. (1986). Specification searches in covariance structure modeling. Psychological Bulletin, 100, 107-120.

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

Marsh, H.W., & Hocevar, D. (1965). Application of confirmatory factor analysis to the study of self-concept: First-and higher older factor models and their invariance across groups. Psychological Bulletin, 97(3), 562-582.

Mulaik, S.A., James, L.R., van Alstein, J., Bennett, N., Lind, S., & Stilwell, C.D. (1989). Evaluations of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Reynolds, W.M., & Gould, J.W. (1981). A psychometric investigation of the standard and short-form of the Beck Depression Inventory. Journal of Consulting and Clinical Psychology, 49, 306-307.

Shaw, B.F., Steer, R.A., Beck, A.T., & Schut, J. (1979). The structure of depression in heroin addicts. British Journal of Addiction, 74, 295-303.

Shek, D.T.L. (1990). Reliability and factorial structure of the Chinese version of the Beck Depression Inventory. Journal of Clinical Psychology, 46, 35-43.

Snook, S.C., & Gorsuch, R.L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. Psychological Bulletin, 106, 148-154.

Steer, R.A., Beck, A.T., Riskind, J.H., & Brown, G. (1987). Relationships between the Beck Depression Inventory and the Hamilton Psychiatric Rating Scale for Depression in depressed outpatients. Journal of Psychopathology and Behavioral Assessment, 9, 327-339.

Steer, R.A., McElroy, M.G., & Beck, A.T. (1982). Structure of depression in alcoholic men: A partial replication. Psychological Reports, 50, 723-728.

Steer, R.A., Shaw, B.F., Beck, A.T., & Fine, E.W. (1977). Structure of depression in black alcoholic men. Psychological Reports, 41, 1235-1241.

Strober, M., Green, J., & Carlson, G. (1981). Utility of the Beck Depression Inventory with psychiatrically hospitalized adolescents. Journal of Consulting and Clinical Psychology, 49, 482-483.

Suyapa, E., Silvia, M., & MacCallum, R.C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. Multivariate Behavioral Research, 23, 297-326.

Tanaka, J.S., & Huba, G.J. (1984). Confirmatory hierarchical factor analyses of psychological distress measures. Journal of Personality and Social Psychology, 46, 621-635.

Tanaka-Matsumi, J., & Kameoka, V.A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. Journal of Consulting and Clinical Psychology, 54, 328-333.

Teri, L, (1982). The use of the Beck Depression Inventory with adolescents. Journal of Abnormal Child Psychology, 10, 227-284.

Vredenburg, K., Krames, L., & Flett, G.L. (1985). Reexamining the Beck Depression Inventory: The long and short of it. Psychological Reports, 56, 767-778.

Walkey, F.W. (1983). Simple versus complex factor analyses of responses to multiple scale questionnaires. Multivariate Behavioral Research, 18, 401-421.

Weckowicz, T.E., Muir, W., & Cropley, A.J. (1967). A factor analysis of the Beck Inventory of Depression. Journal of Consulting Psychology, 31, 23-28.

Appendix

Items Comprising the Beck Depression Inventory[a]

1. I do not feel sad
2. I am not particularly discouraged about the future
3. I do not feel like a failure
4. I get as much satisfaction out of the things as I used to
5. I don't feel particularly guilty
6. I don't feel I am being punished
7. I don't feel disappointed in myself
8. I don't feel I am any worse than anybody else
9. I don't have any thoughts of killing myself
10. I don't cry any more than usual
11. I am no more irritated now than I ever am
12. I have not lost interest in other people
13. I make decisions about as well as I ever could
14. I don't feel I look any worse than I used to
15. I can work about as well as before
16. I can sleep as well as usual
17. I don't get more tired than usual
18. My appetite is no worse than usual

19. I haven't lost much weight, if any lately

20. I am no more worried about my health than usual

21. I have not noticed any recent change in my interest in sex

---

[a] Only the first of four statements for each item is presented here

Footnotes

1. Although the Tanaka and Huba (1984) study also modeled the Beck and Lester (1973) data, the Steer et al. (1977) model was chosen because it was the more parsimonious model of the two and, thus, had the great ŕ probability of replication.
2. An MI may be computed for each constrained parameter and indicates the expected decrease in χ^2 if the parameter were to be relaxed; the decrease, however, may actually be higher. MI values < 5.00 indicate little appreciable improvement in model fit.
3. The same procedure was followed for the validation sample (Group 3).
4. Because χ^2 values are used in the computation of the cross-validation index, these values are reported here for Group 3, but only as they relate to plausible model alternatives.
5. It is important to note that the parameters were relaxed only if it made sense, substantively, to do so.
6. Cross-loadings tend to be very unstable and, consequently, are almost impossible to replicate.
7. The reader will recall that Item #19, measuring weight loss, also produced a problematic loading in the exploratory factor analyses.
8. In contrast to zero-order correlations between pairs of observed variables, these correlations represent associations between latent constructs as measured by multiple observed variables (i.e., sets of items).

Table 1

Exploratory Factor Analytic Results for a 3-Factor Structure (Group 1)[a]

Items		Factor I	Factor II	Factor III
I.	Negative Attitudes			
2.	Pessimism	.65	(.24)	
3.	Failure	.50		
6.	Punishment	.47		
7.	Self-dislike	.43	(.29)	
1.	Sadness	.38	(.25)	
5.	Guilt	.34		
14.	Self-image	.33		
9.	Suicidal	.31		
10.	Crying	.24		
II.	Performance Difficulty			
4.	Dissatisfaction		.65	
17.	Fatigue		.58	
13.	Indecisiveness		.51	
20.	Hypochondria		.44	
16.	Insomnia		.41	
8.	Self-accusation	(.29)	.34	
11.	Irritability		.31	
15.	Work Inhibition	(.22)	.29	
12.	Withdrawal	(.23)	.28	
21.	Libido Loss		.28	
III.	Somatic Elements			
18.	Appetite Loss			1.05[b]
19.	Weight Loss			.18

[a] Cross-loadings ≤ .20 are deleted for sake of clarity; those > .20 are parenthesized

[b] Reflects a multicollinearity problem

Table 2

Confirmatory Factor Analytic Results for 3-Factor First-order Structures

	Group 2 (Calibration)								Group 3 (Validation)				
Competing Model	χ^2	df	Model Comparison	$\Delta\chi^2$	Δdf	AGFI	CFI	PCFI	χ^2	df	AGFI	CFI	PCFI
0 Null Model	1356.89	210	—	—	—	—	—	—	1091.93	210	—	—	—
1 EFA specification	308.98	186	—	—	—	.89	.89	.79	313.09	186	.84	.86	.76
2 Model 1 with Item #8 cross-loading on Factor 1	296.34	185	2 vs 1	12.64***	1	.89	.90	.79	—	—	—	—	—
3 Model 2 with Item #9 cross-loading on Factor 3	285.30	184	3 vs 2	11.04***	1	.89	.91	.80	—	—	—	—	—
4 Model 3 with Item #16 cross-loading on Factor 3	274.43	183	4 vs 3	10.87***	1	.90	.92	.80	—	—	—	—	—
5 Model 4 with Item #21 cross-loading on Factor 3	263.47	182	5 vs 4	10.96***	1	.90	.93	.81	—	—	—	—	—
6 Model 5 with Item #12 cross-loading on Factor 1	256.78	181	6 vs 5	6.69**	1	.90	.93	.80	—	—	—	—	—
7 Model 6 with Items #8, #12, loading on Factor 1	261.98	183	7 vs 1	47.00***	3	.90	.93	.81	289.07	183	.85	.88	.77
8 Model 7 with 3 cross-loadings deleted	296.47	186	8 vs 7	34.49***	3	.89	.90	.80	299.90	186	.84	.86	.75
9 Model 7 with Items #9, #16, #21 loading on Factor 3	301.84	186	8 vs 1	7.14	0[a]	.89	.90	.80	314.93	186	.84	.85	.75
10 Items #8, #12, #9 loading on Factor 1; Items #16, #21 loading on Factor 3	290.22	186	10 vs 1	18.76	0[a]	.89	.91	.81	296.29	186	.85	.87	.77

** $p < .01$ *** $p < .001$

[a] Since this model involves only a shift in factor loadings, and no change in the number of parameters specified, there is no difference in degrees of freedom; thus, no statement can be made regarding statistical significance.

Factor 1 = Negative Attitudes; Factor 2 = Performance Difficulty; Factor 3 = Somatic Elements

Table 3

Confirmatory Factor Analytic Results for 3-Factor Second-order Structures

	Group 2 (Calibration)								Group 3 (Validation)				
Competing Model	χ^2	df	Model Comparison	$\Delta\chi^2$	Δdf	AGFI	CFI	PCFI	χ^2	df	AGFI	CFI	PCFI
0 Null Model	1356.89	210	—	—	—	—	—	—	1091.93	210	—	—	—
1 EFA specification	307.86	186	—	—	—	.89	.89	.79	313.09	186	.84	.86	.76
2 Model 1 with Item #8 cross-loading on Factor 1[a]	295.38	186	2 vs 1	12.48	—	.89	.90	.80	—	—	—	—	—
3 Model 2 with Item #9 cross-loading on Factor 3	284.34	185	3 vs 2	11.04***	1	.90	.91	.80	—	—	—	—	—
4 Model 3 with Item #16 cross-loading on Factor 3	273.45	184	4 vs 3	10.89***	1	.90	.92	.81	—	—	—	—	—
5 Model 4 with Item #21 cross-loading on Factor 3	263.01	183	5 vs 4	10.44**	1	.90	.93	.81	—	—	—	—	—
6 Model 5 with Item #8 loading on Factor 1	265.14	184	6 vs 1	42.72***	2	.90	.93	.81	291.28	184	.84	.88	.77
7 Model 6 with 3 cross-loadings deleted	297.31	187	7 vs 6	32.17***	3	.89	.90	.80	313.49	187	.84	.87	.77
8 Model 6 with Items #9, #16, #21 loading on Factor 3	297.46	187	7 vs 1	10.40**	1	.89	.90	.80	304.31	187	.84	.86	.77
9 Items #8, #9 loading on Factor 1; Items #16, #21 loading on Factor 3	290.45	187	9 vs 1	17.41***	1	.89	.91	.81	303.90	187	.84	.87	.77
10 Tanaka & Huba 2nd-order structure[b]	322.35	186	10 vs 9	31.90***	1	.88	.88	.78	319.73	186	.83	.85	.75

** $p < .01$ *** $p < .001$

[a] Variance estimate for Factor 2 was fixed to .001 to offset presence of a Heywood case, thus accounting for the additional degree of freedom

[b] Tanaka & Huba (1984); Figure 2 based on Steer et al. (1977) data

Factor 1 = Negative Attitudes; Factor 2 = Performance Difficulty; Factor 3 = Somatic Elements

Table 4

Double Cross-validation Results Across Calibration and Validation Samples

	Group 2 (Calibration)[a] (N = 279)			Group 3 (Validation)[b] (N = 189)		
Competing Model	χ^2	df	CVI[c]	χ^2	df	CVI
First-order Structures						
1 EFA specification	382.92	210	1.38	363.46	210	1.93
7 Model 1 with Items #8, #12 loading on Factor 1; Items #9, #16, #21 cross-loading on Factor 3	374.48	210	1.35	347.92	210	1.85
8 Model 7 with 3 cross-loadings deleted	395.53	210	1.42	362.70	210	1.93
9 Model 7 with Items #9, #16, #21 loading on Factor 3	369.58	210	$\boxed{1.33}$	356.72	210	1.90
10 Items #8, #12, #9 loading on Factor 1; Items #16, #21 loading on Factor 3	379.81	210	1.37	359.40	210	1.91
Second-order Structures						
1 EFA specification	376.20	207	1.35	360.53	208	1.92
6 Model 1 with Item #8 loading on Factor 1; Items #9, #16, #21 cross-loading on Factor 3	489.87	208	1.76	343.52	208	1.83
7 Model 6 with 3 cross-loadings deleted	360.70	208	1.30	350.16	208	1.86
8 Model 6 with Items #9, #16, #21 loading on Factor 3	366.94	208	1.32	361.22	208	1.92
9 Items #8, #9 loading on Factor 1; Items #16, #21 loading on Factor 3	349.31	208	$\boxed{1.26}$	347.31	208	1.85
10 Tanaka & Huba[c] Model	401.66	207	1.44	371.78	207	1.98

[a] Model estimations based on Group 2 covariance matrix; model specification constrained factor loading and correlation parameters to be fixed a priori based on Group 3 estimates

[b] Model estimation based on Group 3 covariance matrix; model specification constrained factor loading and correlation parameters to be fixed a priori based on Group 2 estimates

[c] CVI = χ^2 / N - 1

[d] Tanaka & Huba (1984) - Figure 2 based on Steer et al. (1977) data

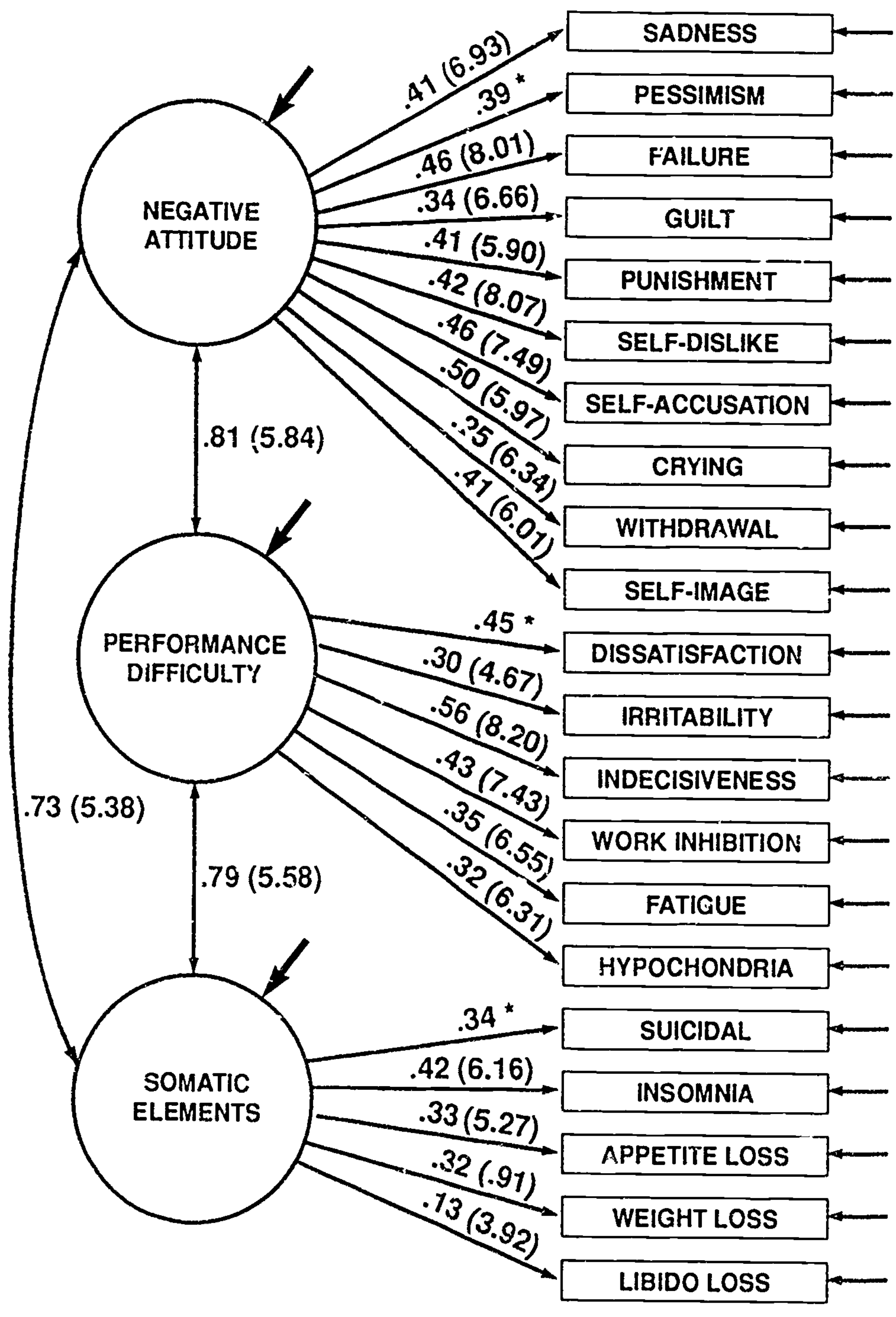
Figure Captions

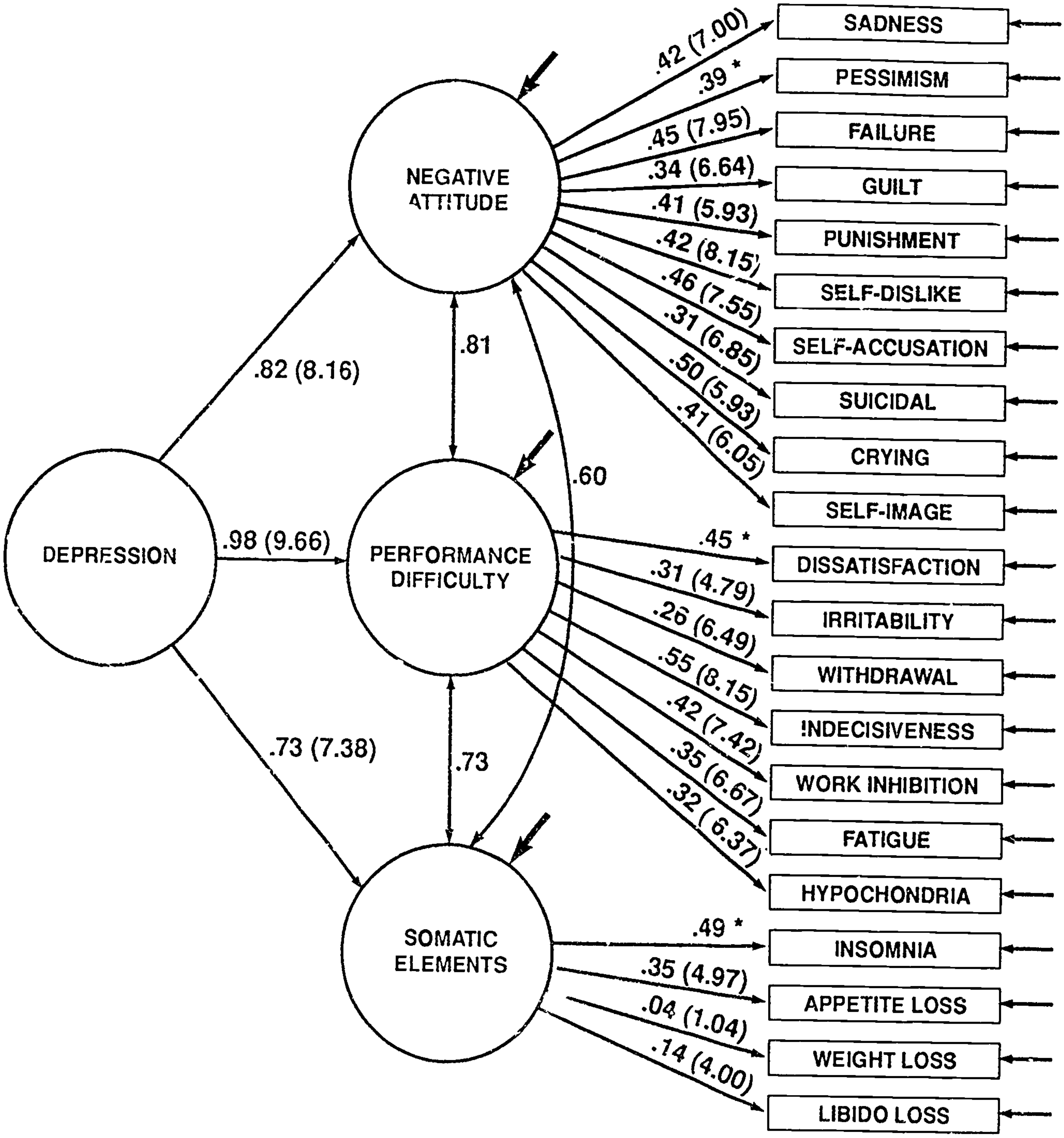
Figure 1 Standardized estimates for final (Model 7) first-order 3-factor structure underlying the Beck Depression Inventory based on Group 2 (calibration sample). (Critical ratios are parenthesized; values > 1.96 indicate statistical significance.)

* denotes parameter fixed to 1.0 in the original solution for purposes of statistical identification.

Figure 2 Standardized estimates for final (Model 9) second-order 3-factor structure underlying the Beck Depression Inventory based on Group 2 (calibration sample). (Critical ratios of estimates are parenthesized; values > 1.96 indicate statistical significance.)

* denotes parameter fixed to 1.0 in the original solution for purposes of statistical identification


NEGATIVE ATTITUDE
PERFORMANCE DIFFICULTY
SOMATIC ELEMENTS
.81 (5.84)
.79 (5.58)
.73 (5.38)
.41 (6.93)
.39 *
.46 (8.01)
.34 (6.66)
.41 (5.90)
.42 (8.07)
.46 (7.49)
.50 (5.97)
.35 (6.34)
.41 (6.01)
.45 *
.30 (4.67)
.56 (8.20)
.43 (7.43)
.35 (6.55)
.32 (6.31)
.34 *
.42 (6.16)
.33 (5.27)
.32 (.91)
.13 (3.92)
SADNESS
PESSIMISM
FAILURE
GUILT
PUNISHMENT
SELF-DISLIKE
SELF-ACCUSATION
CRYING
WITHDRAWAL
SELF-IMAGE
DISSATISFACTION
IRRITABILITY
INDECISIVENESS
WORK INHIBITION
FATIGUE
HYPOCHONDRIA
SUICIDAL
INSOMNIA
APPETITE LOSS
WEIGHT LOSS
LIBIDO LOSS



DEPRESSION
NEGATIVE ATTITUDE
PERFORMANCE DIFFICULTY
SOMATIC ELEMENTS
.82 (8.16)
.98 (9.66)
.73 (7.38)
.81
.73
.60
.42 (7.00)
.39 *
.45 (7.95)
.34 (6.64)
.41 (5.93)
.42 (8.15)
.46 (7.55)
.31 (6.85)
.50 (5.93)
.41 (6.05)
.45 *
.31 (4.79)
.26 (6.49)
.55 (8.15)
.42 (7.42)
.35 (6.67)
.32 (6.37)
.49 *
.35 (4.97)
.04 (1.04)
.14 (4.00)
SADNESS
PESSIMISM
FAILURE
GUILT
PUNISHMENT
SELF-DISLIKE
SELF-ACCUSATION
SUICIDAL
CRYING
SELF-IMAGE
DISSATISFACTION
IRRITABILITY
WITHDRAWAL
INDECISIVENESS
WORK INHIBITION
FATIGUE
HYPOCHONDRIA
INSOMNIA
APPETITE LOSS
WEIGHT LOSS
LIBIDO LOSS