DOCUMENT RESUME

ED 335 406 TM 017 089

AUTHOR Byrne, Barbara M.; And Others
TITLE The Beck Depression Inventory: Testing for Invariant Measurement and Structure across Gender for Nonclinical Adolescents.
PUB DATE 91
NOTE 31p.; Paper presented at the Annual Meeting of the American Psychological Association (San Francisco, CA, August 16-20, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adolescents; *Affective Measures; Analysis of Covariance; Comparative Testing; Factor Structure; Females; Foreign Countries; Item Analysis; Males; Secondary Education; *Secondary School Students; *Sex Differences; *Test Items
IDENTIFIERS *Beck Depression Inventory; Canada; Covariance Structural Analysis; *Invariance Principle

ABSTRACT

Extending the earlier work of B. M. Byrne and P. Baron (1990), the factorial invariance of the 21-item Beck Depression Inventory (BDI) was tested using 351 non-clinical adolescent males and 334 non-clinical adolescent females. All subjects were in grades 9 through 12 and attended the same secondary school in a large metropolitan area in central Canada. Participants, who ranged in age from 12 to 18 years, completed the BDI and other assessment measures during one regular class period. All analyses were based on confirmatory factor analytic procedures within the framework of covariance structure modeling. Despite the differential loading pattern of Item 20 (hypochondria) for females, a three-factor structure that comprised one second-order general factor of depression and three first-order factors representing negative attitudes, performance difficulty, and somatic elements was found to be invariant across gender. Furthermore, except for two additional inventory items (items 8 and 10, self-accusation and crying, respectively), the remaining BDI measurements were group-invariant. The results are expected to be of substantial interest to both researchers and clinicians whose concerns focus on depression as it affects adolescents. A 40-item list of references is included. Two data tables and two figures are provided. Items comprising the BDI are appended. (RLC)

The Beck Depression Inventory: Testing for Invariant Measurement and Structure Across Gender for Nonclinical Adolescents

Barbara M. Byrne

T. Leanne Campbell   Pierre Baron

University of Ottawa

Paper presented at the American Psychological Association Annual Meeting, San Francisco 1991

Abstract

Extending the earlier work of Byrne and Baron (1990), the factorial invariance of the Beck Depression Inventory (BDI) was tested across nonclinical adolescent males (n=351) and females (n=334) using confirmatory factor analytic procedures. Despite the differential loading pattern of Item 20 for females, a 3-factor structure that comprised one 2nd-order general factor of depression and three 1st-order factors representing negative attitudes, performance difficulty, and somatic elements was found to be invariant across gender. Furthermore, except for two additional items (#8, #10), the remaining BDI measurements were group-invariant. Results are expected to be of substantial interest to both researchers and clinicians whose concerns focus on depression as it bears on this population.

Measuring Adolescent Depression: Factorial Validity and Invarance of the Beck Depression Inventory Across Gender

Although the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) was originally developed for use with clinical populations, a review of the literature reveals its popularity as a measuring instrument for nonclinical populations as well. In particular, the BDI has been used almost exclusively in the assessment of depression for nonclinical adolescents. Cross-validated findings from a recent factor analytic study of the BDI for this population (Byrne & Baron, 1990) demonstrated strong support for an hierarchical structure consisting of one higher-order general factor of depression, and three lower-order factors representing Negative Attitudes, Performance Difficulty, and Somatic Elements. The present study extends the work of Byrne and Baron by testing for the equivalence of BDI item measurements and a second-order factorial structure across adolescent males and females.

A frequent finding in epidemiological studies of depression is that adult females report a significantly higher rate of depression than males; a ratio of 2 to 1 has typically been found for both clinical and nonclinical samples (Hirschfield & Cross, 1982; Nolen-Hoeksema, 1987; Weissman & Klerman, 1977). Indeed, similar findings have recently been reported for adolescents (Allgood-Merten, Lewinsohn, & Hops, 1990; Baron & Perron, 1986; Campbell, Byrne, & Baron, 1990; Reynolds, 1986; Rutter, Izard, & Read, 1986).

Important assumptions in testing for gender differences in depression scores based on the BDI, however, are that (a) all item scaling units are equivalent across gender (i.e., all items are interpreted in exactly the same way by males and females), and (b) the factorial structure is equivalent across males and females (i.e., relations among the facets of depression remain consistent across gender). Despite the fact that violation of these assumptions clearly invalidates interpretations based on mean score differences, no study to date has tested the validity of these equivalencies.

In this regard it may be, as demonstrated in recent construct validity research related to adolescent self-concept, (Byrne, 1990; Byrne & Shavelson, 1987), that interrelations among the facets of depression vary somewhat for adolescent males and females. Furthermore, it is possible that a different pattern of items defines the underlying factorial structure of the BDI for each sex. Such findings would suggest either a differential perception of item content, or a differential nomological network of depression structure across gender. These questions bear on the construct validity of adolescent depression and can be formulated into hypotheses that can be tested statistically. This was the task of the present study, as it related to the factorial structure of the BDI proposed by Byrne and Baron (1990).

Building upon the earlier work of Tanaka and Huba (1984), Byrne and Baron (1990) used confirmatory factor analyses (CFAs) within the framework of covariance structure modeling to test the validity of a 2nd-order structure. In contrast to Tanaka and Huba, whose validity work was based on clinical adult data, Byrne and Baron used three independent samples of nonclinical adolescents (mean age=15 yrs) to test for, and cross-validate the factorial validity of the BDI. Consistent with the Tanaka and Huba findings, Byrne and Baron found an hierarchical factorial structure to most adequately represent the data, albeit certain item loadings differed for the clinical and nonclinical populations. On the basis of their cross-validated findings, Byrne and Baron concluded that the underlying structure of BDI responses comprises one 2nd-order general factor of depression, and three 1st-order factors which they labelled Negative Attitudes, Performance Difficulty, and Somatic Elements. This model is presented schematically in Figure 1.

Insert Figure 1 about here

Except for the work of Byrne and Baron (1990), all previous factor analyses of the BDI, as they relate to the nonclinical adolescent population, have used exploratory (EFA) procedures involving principal components analyses with varimax rotation. Findings from this research, however, have been inconsistent; 2-factor (Shek, 1990), 3-factor (Baron & LaPlante, 1984), and 4-factor (Teri, 1982) solutions have been reported. This discordance can be linked to the widely known limitations associated with EFA procedures in general (e.g., Bollen, 1989; Long, 1983), and principal components analyses in particular (e.g., Borgatta, Kercher, & Stull, 1986; Gorsuch, 1990; Hubbard & Allen, 1987; Snook & Gorsuch, 1989).

The present study addressed the methodological weaknesses of previous factor analytic and gender difference research by assessing the equivalency of BDI measurement and structure across adolescent males and females based on a rigorously tested and cross-validated factorial model. All analyses were based on CFA procedures within the framework of covariance structure modeling.

Method

Sample and Procedures

The data comprised BDI responses from 685 (males, n=351; females, n=334) adolescents (grades 9-12) attending the same secondary school in a large metropolitan area in central Canada. Participants ranged in age from 12 to 18 years (M=15 years). Only questionnaires with complete data were included in the analyses.

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 ques ions were solicited and answered. All participation, in keeping with school and Ethics Committee policies, was voluntary and no incentives were offered.

Instrumentation

The BDI is a 21-item scale that measures symptoms 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, Steer, & Garbin, 1988; Kearns, Cruickshaw, McGuigan, Riley, Shaw, & Snaith, 1982; Vredenburg, Krames, & Flett, 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; higher scores represent a more severe level of 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). Zero-rated statements for BDI items are presented in the Appendix.

Previous research has demonstrated the BDI to be psychometrically sound in measuring depression for nonclinical adolescents. Tests for the internal consistency reliability of the BDI for this population have revealed alpha coefficients ranging from .80 to .90 (mean α =.86; Baron & Laplante, 1984; Barrera & Garrison-Jones, 1988; Shek, 1990; Teri, 1982). A test-retest reliability coefficient of .74 has also been reported (Baron & LaPlante, 1984). Additionally, Barrera and Garrison-Jones have demonstrated evidence of convergent validity with the Child Assessment Schedule (CAS; Hodges, Kline, Stern, Cytryn, & McKnew, 1982) for items measuring depression symptoms (r=.73) and with the General Self-worth subscale of the Perceived Competence Scale for Children (Harter, 1982; r=-.64); they also reported evidence of discriminant validity with CAS items measuring conduct disorder (r=.29) and anxiety (r=.29) symptoms.

Data Analyses

Data were analyzed in two stages. First, the 3-factor 2nd-order structure proposed by Byrne and Baron (1990) was tested separately for goodness-of-fit to male and female data. Because it is important that the model for each sex be well-fitting before testing for equivalencies of measurement and structure across the two groups, model respecification was conducted if there was statistical and theoretical justification for doing so.[1] Once the baseline model was determined for each sex, the equivalence of item measurements and factorial structure was tested across males and females.

Cognizant of caveats related to tests for partial measurement procedures (see Byrne, Shavelson, & Muthén, 1989), equality constraints were imposed on only those parameters similarly specified for both groups.

Assessment of model fit was based on multiple criteria that reflected statistical, theoretical, and practical considerations; these included (a) the X^2 likelihood ratio, (b) the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), (c) the relative noncentrality index (RNI; McDonald & Marsh, 1990), (d) the adjusted goodness-of-fit index (AGFI), t-values (parameter estimates relative to their standard errors of estimate), and modification indices (MIs)[2], all provided by the LISREL VI computer program (Joreskog & Sorbom, 1985), and (e) the substantive meaningfulness of the model (see MacCallum, 1986; Suyapa, Silvia, & MacCallum, 1988). Criteria indicative of adequate model fit were values >.90 for the TLI and RNI, and >2.00 for the t-values.

Results

Determination of Baseline Models

For adolescent males, as indicated in Table 1, the hypothesized 3-factor hierarchical model of depression represented a reasonably good fit to the data; all parameters were statistically significant. (Model 0 argues that each item represents a factor, and provides the null model required for computation of the TLI and RNI.) A review of the MIs, however, indicated that model respecification could yield a significantly better fit if Item 12 (loss of interest in other people) were free to cross-load onto the Somatic Elements factor (MI=16.75); reestimation yielded a statistically nonsignificant cross-loading (t=.429). More importantly, however, we could provide no theoretical rationale for this parameterization. Thus, the originally hypothesized factorial model was considered optimal in representing the BDI for adolescent males.

Insert Table 1 about here

For adolescent females, on the other hand, model fit for the initial parameterization was somewhat less well-fitting (see Table 1). Furthermore, one item (#19, measuring weight loss) was found to be nonsignificant (t=.690), and a review of the MIs revealed two parameters to be worthy of estimation. The more prominent fixed parameter (MI=24.05) represented an error correlation between Item 20 (concern for health) and Item 21 (interest in sex), while the other (MI=12.65) represented the cross-loading of Item 20 onto the Negative Attitudes factor. Three additional models were therefore specified and estimated; as shown in Table 1, all yielded highly significant improvements in model fit.[3] The first of these post hoc models (Model 2) specified an error correlation between Items 20 and

21, the second (Model 3) specified, additionally, that Item 20 cross-load onto the Negative Attitudes factor, and the third (Model 4), that Item 20 load on the Negative Attitudes factor, rather than on the Performance Difficulty factor. As shown in Table 1, goodness-of-fit for Models 3 and 4 were almost identical; given these results, parsimony was the key element in choosing between Models 3 and 4. For statistical, psychometric, and theoretical reasons, then, Model 4 was considered to be the most plausible in representing BDI data for adolescent females. Baseline models representing the data for males and for females are presented schematically in Figure 2.

Insert Figure 2 about here

Determination of Factorial Invariance

Tests for invariance were carried out in two stages. First, all item measurements were testea for their invariance across males and females. Once it was determined which items were operating equivalently across groups, these items were constrained equal and a 2nd-order structure linking the three 1st-order factors to general depression was tested next for its invariance across gender. (For a more detailed discussion and description of CFA applications using LISREL modeling, see Byrne, 1989.)

These analyses involved first specifying a model in which all 1st-order factor loadings were constrained equal across samples, and then comparing that model with a less restrictive one in which these parameters were free to take on any value. As with model testing noted earlier, the difference in fit between the two nested models provides a basis for determining the tenability of the hypothesized equality constraints; a significant X^2 indicates noninvariance, and thus rejection of the hypothesis. Turning to Table 2, we see that the comparison of Model 2 (in which these parameters were constrained equal) with Model 1 (in which they were free to vary), yielded a highly significant ΔX^2 (37.24, p<.001), thereby substantiating rejection of the hypothesis of invariant item measurements across males and females.

Insert Table 2 about here

Since this result indicated that certain BDI items were gender-specific in their measurement of lower-order facets of depression, the next task was to identify these noninvariant items; this required the specification of six additional models. Models 3, 4, and 5 tested separately for the invariance of items measuring Negative Attitudes, Performance Difficulty, and Somatic Elements, respectively; all items comprising the latter two factors were found to be equivalent across gender.

Thus, to pinpoint which items related to the Negative Attitudes factor were contributing to the inequality of measurement, tests for invariance were conducted separately for each item. (Because Item 20 was known to operate differently for males and females, it was never constrained equal across groups; see Byrne et al., 1989.) As indicated in Table 2, these analyses revealed Items 8 and 10 not to be equivalent across gender.

Having identified the items that were operating equivalently for males and females, we next specified a model in which these 1st-order loadings, and the three 2nd-order loadings were constrained equal (Model 7). This model was compared with Model 6 in which the 2nd-order loadings were unconstrained. The comparison yielded findings that were not significant ($\Delta\chi^2(3)=6.03$, $p>.05$), thereby supporting the equivalence of an hierarchically-ordered factorial structure across adolescent males and females.

Discussion

In testing first for the validity of a 3-factor 2nd-order structure of depression for adolescent males and females, as initially hypothesized by Byrne and Baron (1990), we found a differential factor loading pattern related to Item 20 (hypochondria) for females; additionally, a substantial correlated error between Items 20 and 21 (libido loss) was evidenced only for females. That these findings should be present for adolescent females, and not for adolescent males, is both interesting and substantively reasonable. Error correlations between item-pairs are usually an indication of perceived redundancy in item content. For adolescent females, then, it appears that Items 20 and 21 elicit responses reflective of the same mental set. Considering the degree of social attention accorded sexually-transmitted diseases in general, and AIDS in particular in recent years, it is not surprising that young people develop health concerns relative to sexual activity. Of particular relevance is the fact that caveats conveyed by various media have focused more on young females than on young males.

Relatedly, Rierdan, Koff, and Stubbs (1987, 1988) have noted that adolescent females tend to exhibit lower levels of body self-image than their male counterparts. This finding may well explain the loading of Item 20 on the Negative Attitudes factor, rather than on the Performance Difficulty factor, as was the case for males. Another possible explanation of this differential loading pattern might be that health complaints are more stongly linked to performance difficulty for males than for females.

Tests for invariance revealed that Item 8 (self-accusation) and Item 10 (crying) were not measuring the factor of Negative Attitudes equivalently for adolescent males and females. One possible explanation for the nonequivalency of Item 8 might lie with differences in depressive experiences across gender. For example, Blatt and Shichman (1986) have identified two types of depressive experiences -- self-critical and dependent; research has shown the self-critical style to be unrelated to femininity (Welkowitz, Lish, & Bond, 1985). Moreover, these findings may also explain why the factor loading is much higher for males (.55) than for females (.38).

The differential measurement of Negative Attitudes for Item 10 could be linked to social reinforcement factors. Because it is more socially acceptable for females to cry than males, it may well be that females are therefore more apt to express symptoms of depression (i.e., manifested in negative attitudes) in this manner. Consistent with socially prescribed gender roles, the loading of crying onto the Negative Attitudes factor is higher for females (.48), than it is for males (.30).

One limitation of the present study was the use of maximum likelihood (ML) estimation procedures, in light of some nonnormality in the data for both males and females. Although the estimates were likely unaffected, it is possible that the t-values were inflated as a consequence of downwardly biased standard errors (Muthen & Kaplan, 1985; Sharma, Durvasula & Dillon, 1989). However, since the hypothesized model of BDI factorial structure used in the present study was based on the Byrne and Baron (1990) parameterization as derived from ML estimation, we considered it important to maintain consistency with the previous CFA study. Analyses based on asymptotic, rather than on normal distributional properties, may yield a slightly different factorial structure. However, further factor analytic research bearing on use of the BDI with nonclinical adolescents is necessary to examine this possibility.

Conclusion

Although results of the study confirmed the hypothesized hierarchical structure of the BDI and determined the equality of theoretical relations among the three facets of depression for adolescent males and females, three measurement-related differences were found: (a) Item 20 was found to load on the Negative Attitudes factor, rather than on the Performance Difficulty factor for females only, and (b) Items 8 and 10 were found to be nonequivalent in their measurement of Negative Attitudes across gender. However, the fact that our rigorous testing of the BDI yielded only three gender-specific differences is truly remarkable, and speaks well for use of the BDI with nonclinical adolescents. These results should be of substantial interest to researchers and practitioners alike. Indeed, our findings should bear importantly on future research that focuses on methodological issues related to use of the BDI with nonclinical adolescents, and on substantive issues involving multigroup comparisons across gender. From an applied perspective, findings carry important implications both for the application of the BDI, and for the interpretation of its test scores relative to the nonclinical adolescent population. In particular, practitioners are urged to be cognizant of the gender specificity associated with Item 8 (self-accusation), Item 10 (crying), and Item 20 (hypochondria).

References

Allgood-Merten, B., Lewinsohn, P.M., & Hops, H. (1990). Sex differences and adolescent depression. Journal of Abnormal Psychology, 99, 55-63.

Baron, P., & Perron, L.M. (1986). Sex differences in the Beck Depression Inventory scores of adolescents. Journal of Youth and Adolescence, 15, 165-171.

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.

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., 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.

Blatt, S.J. & Schichman, S. (1986). Two primary configurations of psychopathology. Psychoanalysis and Contemporary Thought, 6, 187-254.

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

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.

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

Byrne, B.M. (1990). Investigating sex differences in adolescent self-concept: A look beneath the surface. Applied Measurement in Education, 3, 255-274.

Byrne, B.M., & Baron, P. (1990). The Beck Depression Inventory: A cross-validated study of factorial structure for nonclinical adolescents. ERIC Document Reproduction Services No. TM015499.

Byrne, B.M., & Shavelson, R.J. (1987). Adolescent self-concept: Testing the assumption of equivalent structure across gender, American Educational Research Journal, 24, 365-385.

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, T.L., Byrne, B.M., & Baron, P. (1990, November). Patterns of depressive responses in early adolescents. Paper presented at the Annual Meeting of the Quebec Society for Research in Psychology, Montreal.

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

Harter, S. (1982). The Perceived Competence Scale for Children. Child Development, 53, 87-97.

Hirschfeld, R.M., & Cross, C.K. (1982). Epidemiology of affective disorders. Archives of General Psychiatry, 39, 35-46.

Hodges, K., Kline, J., Stern, L., Cytryn, L., & McKnew, D. (1982). The development of a child assessment interview for research and clinical use. Journal of Abnormal Child Psychology, 10, 173-189.

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., & Sorböm, D. (1985). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood. Mooresville, Ind.: Scientific Software.

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.

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.

McDonald, R.P. & Marsh, H.W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. Psychological Bulletin, 107, 247-255.

Muthén, B., & Kaplan, D. (1985). A comparison of methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depressive symptom expression and help-seeking among college students. Sex Roles, 7, 309-320.

Reynolds, W.M. (1986). About myself RADS form. Florida: Psychological Assessment Resources.

Rierdan, J., Koff, E., & Stubbs, M.L. (1987). Depressive symptomatology and body image in adolescent girls. Journal of Early Adolescence, 7, 205-216.

Rierdan, J., Koff, E., & Stubbs, M.L. (1988). Gender, depression, and body image in early adolescents. Journal of Early Adolescence, 8, 109-117.

Sharma, S., Durvasula, S., & Dillon, W.R. (1989). Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. Journal of Marketing Research, 26, 214-221.

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.

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.

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

Tucker, L.R. & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

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

Weissman, M.M., & Klerman, G.L. (1977). Sex differences and the epidemiology of depression. Archives of General Psychiatry, 34, 98-111.

Welkowitz, J., Lish, J.D., & Bond, R.N. (1985). The Depressive Experiences Questionnaire: Revision and validation. Journal of Personality Assessment, 49, 89-94.

Appendix

<u>Items Comprising the Beck Depression Inventory</u>[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. A baseline model is the most parsimonious, albeit best-fitting and most substantively meaningful model to represent the observed data.
2. An MI can 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. Only MI values > 5.00 were considered since smaller values indicate little appreciable improvement in model fit (see Joreskog & Sorbom, 1985).
3. Nested models can be compared with one another by computing the difference in their χ^2 values and degrees of freedom; this χ^2 difference is itself χ^2- distributed with degrees of freedom equal to the difference in degrees of freedom.

Table 1

Summary of Model Fit Statistics for Males and Females

Model	χ^2	df	AGFI	TLI	RNI
Males					
0 Null Model	1679.73	210	---	---	---
1 Hypothesized Model	316.77	187	.91	.90	.91
Females					
0 Null Model	1592.28	210	---	---	---
1 Hypothesized Model	355.87	187	.89	.86	.88
2 Error correlation between items 20/21	331.13	186	.89	.88	.90
3 Item 20 cross-loaded on F_1	314.26	185	.90	.89	.91
4 Item 20 loaded on F_1, instead of F_2	315.45	186	.90	.89	.91

F_1 = Factor 1 (Negative Attitudes)

F_2 = Factor 2 (Performance Difficulty)

Table 2

Summary of Tests for Invariance Across Males and Females

Model	χ^2	df	Model Comparison	$\Delta\chi^2$	Δdf
1 Baseline Multigroup model	632.22	373	---	---	---
2 Pattern of item loadings invariant[a]	669.46	390	2 vs 1	37.24***	17
3 Item loadings for F_1 invariant	655.96	382	3 vs 1	19.74**	9
4 Item loadings for F_2 invariant	639.45	378	4 vs 1	7.23	5
5 Item loadings for F_3 invariant	638.49	376	5 vs 1	6.27	3
6 All item loadings except #8 and #10 invariant[b]	654.31	388	6 vs 1	22.09	15
7 Model 6 with all 2nd-order loadings invariant	660.34	391	7 vs 6	6.03	3

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

[a] Item #20 was not constrained equal across gender.

[b] Equality constraints were imposed separately for each item loading.

$\Delta\chi^2$ = difference in χ^2 values; Δdf = difference in degrees of freedom.

F_1 = Factor 1 (Negative Attitudes)

F_2 = Factor 2 (Performance Difficulty)

F_3 = Factor 3 (Somatic Elements)

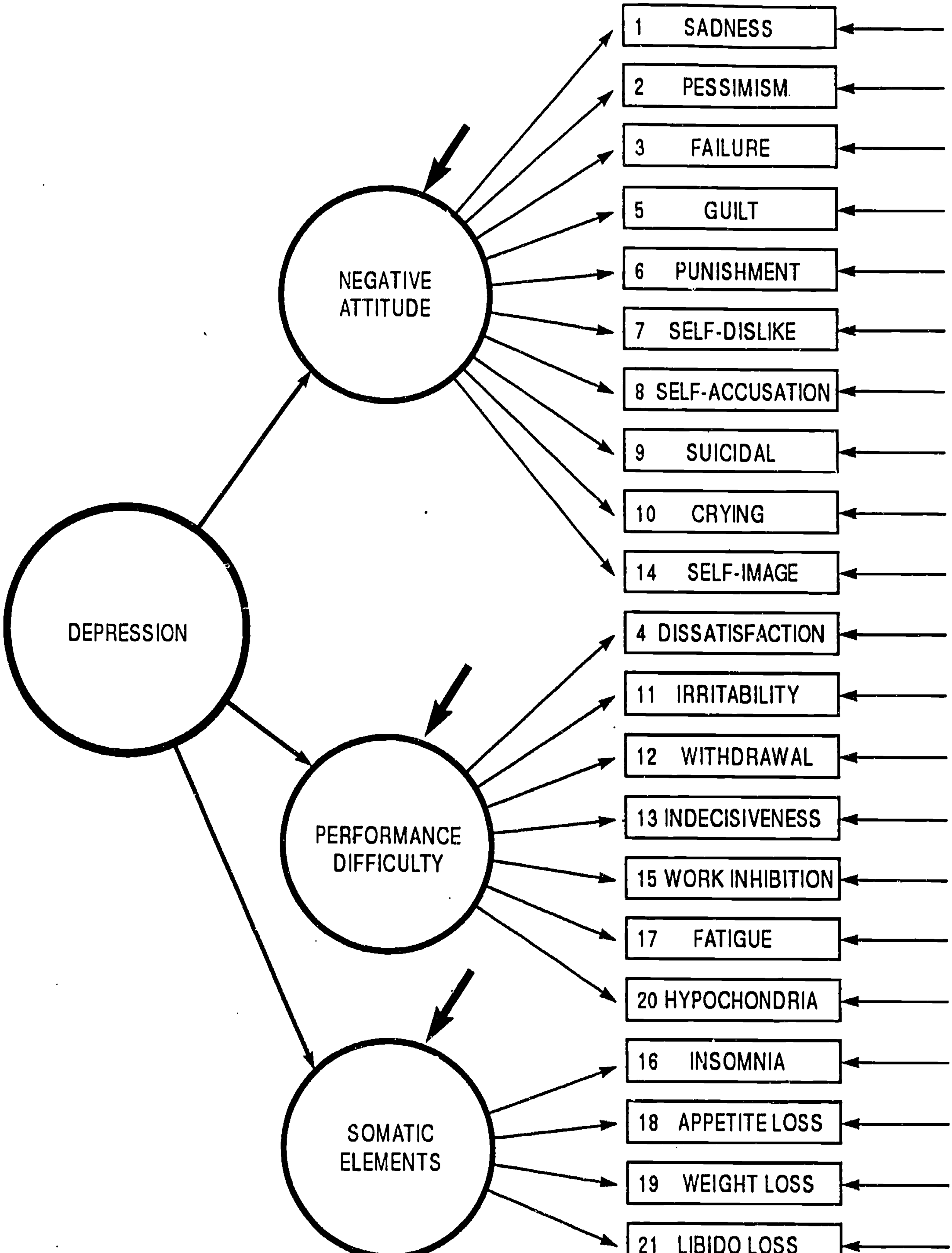
Figure Captions

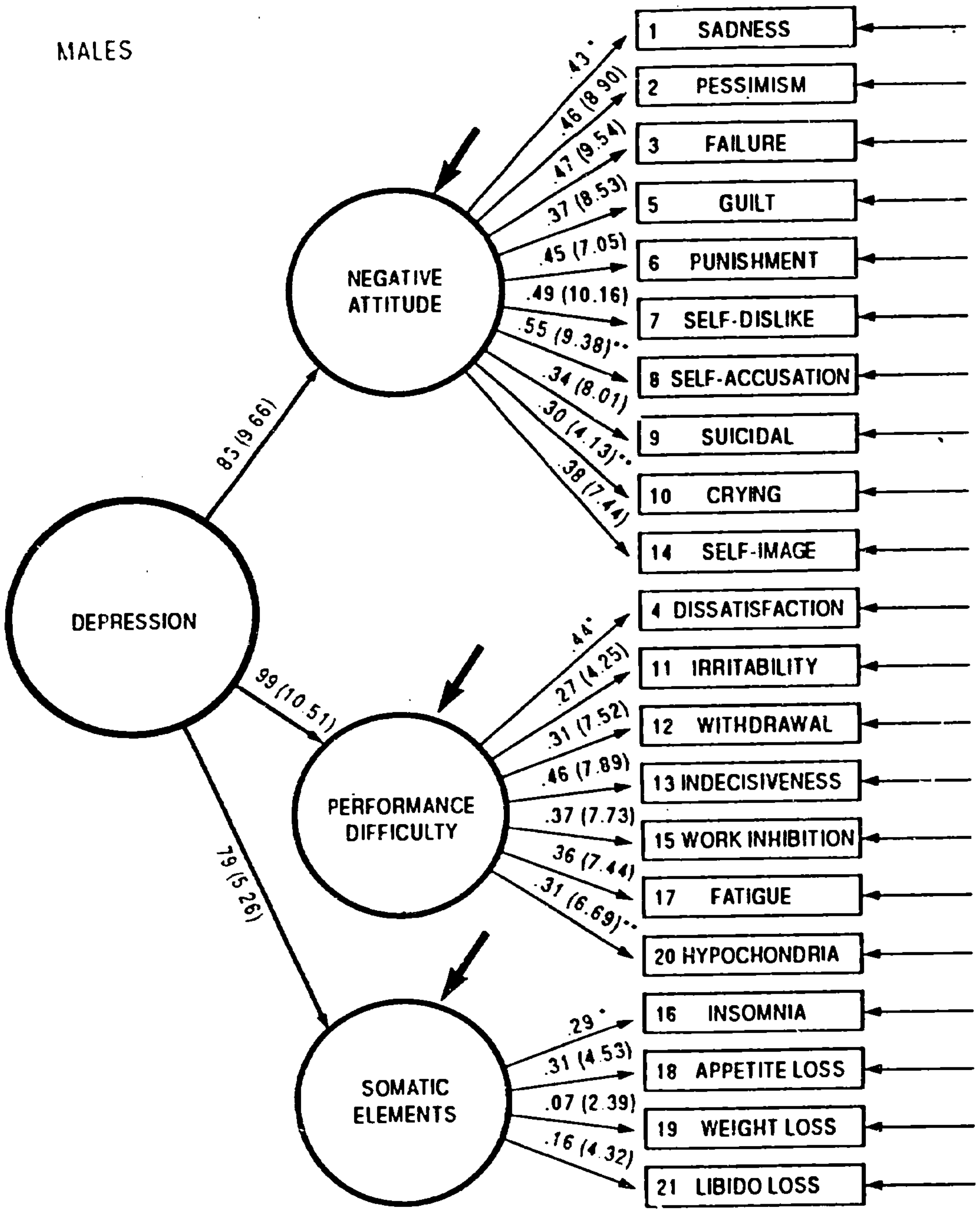
Figure 1. Hypothesized 2nd-order Factorial Structure of the Beck Depression Inventory for Nonclinical Adolescents (Byrne & Baron, 1990).

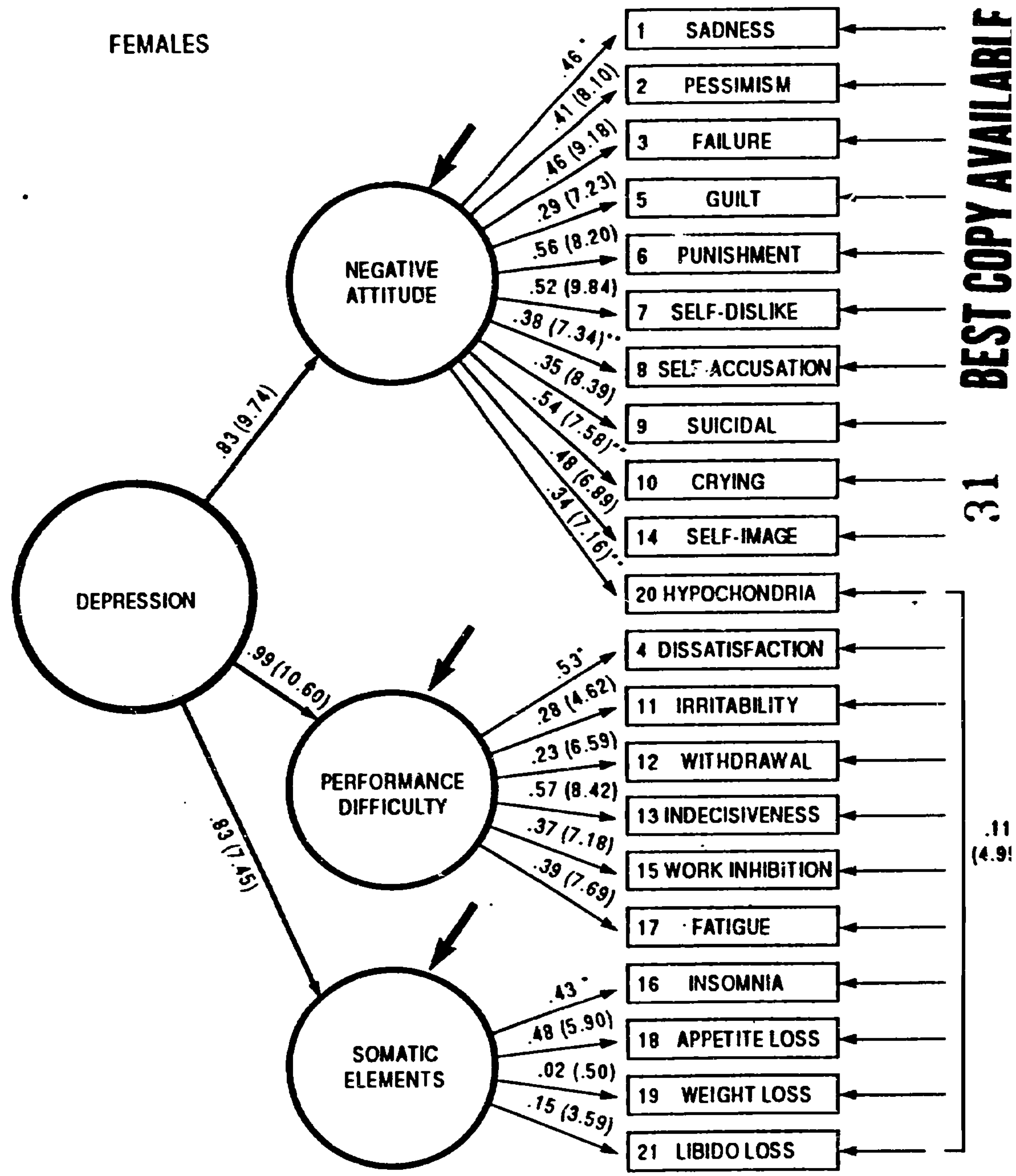
Figure 2. Standardized Estimates for 2nd-order Factorial Structure of the Beck Depression Inventory for Nonclinical Adolescent Males and Females. Parenthesized values represent critical ratios of estimates; values > 1.96 indicate statistical significance ($p < .05$). Values in boxes represent item numbers.

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

** denotes noninvariance across gender.


DEPRESSION
NEGATIVE ATTITUDE
PERFORMANCE DIFFICULTY
SOMATIC ELEMENTS
1 SADNESS
2 PESSIMISM
3 FAILURE
5 GUILT
6 PUNISHMENT
7 SELF-DISLIKE
8 SELF-ACCUSATION
9 SUICIDAL
10 CRYING
14 SELF-IMAGE
4 DISSATISFACTION
11 IRRITABILITY
12 WITHDRAWAL
13 INDECISIVENESS
15 WORK INHIBITION
17 FATIGUE
20 HYPOCHONDRIA
16 INSOMNIA
18 APPETITE LOSS
19 WEIGHT LOSS
21 LIBIDO LOSS


MALES
DEPRESSION
NEGATIVE ATTITUDE
PERFORMANCE DIFFICULTY
SOMATIC ELEMENTS
.83 (9.66)
.99 (10.51)
.79 (5.26)
.43 *
.46 (8.90)
.47 (9.54)
.37 (8.53)
.45 (7.05)
.49 (10.16)
.55 (9.38)**
.34 (8.01)
.30 (4.13)**
.38 (7.44)
.44*
.27 (4.25)
.31 (7.52)
.46 (7.89)
.37 (7.73)
.36 (7.44)
.31 (6.69)**
.29 *
.31 (4.53)
.07 (2.39)
.16 (4.32)
1 SADNESS
2 PESSIMISM
3 FAILURE
5 GUILT
6 PUNISHMENT
7 SELF-DISLIKE
8 SELF-ACCUSATION
9 SUICIDAL
10 CRYING
14 SELF-IMAGE
4 DISSATISFACTION
11 IRRITABILITY
12 WITHDRAWAL
13 INDECISIVENESS
15 WORK INHIBITION
17 FATIGUE
20 HYPOCHONDRIA
16 INSOMNIA
18 APPETITE LOSS
19 WEIGHT LOSS
21 LIBIDO LOSS


FEMALES
DEPRESSION
NEGATIVE ATTITUDE
PERFORMANCE DIFFICULTY
SOMATIC ELEMENTS
.83 (9.74)
.99 (10.60)
.83 (7.45)
.46 *
.41 (8.10)
.46 (9.18)
.29 (7.23)
.56 (8.20)
.52 (9.84)
.38 (7.34)**
.35 (8.39)
.54 (7.58)**
.48 (6.89)
.34 (7.16)**
.53*
.28 (4.62)
.23 (6.59)
.57 (8.42)
.37 (7.18)
.39 (7.69)
.43 *
.48 (5.90)
.02 (.50)
.15 (3.59)
1 SADNESS
2 PESSIMISM
3 FAILURE
5 GUILT
6 PUNISHMENT
7 SELF-DISLIKE
8 SELF-ACCUSATION
9 SUICIDAL
10 CRYING
14 SELF-IMAGE
20 HYPOCHONDRIA
4 DISSATISFACTION
11 IRRITABILITY
12 WITHDRAWAL
13 INDECISIVENESS
15 WORK INHIBITION
17 FATIGUE
16 INSOMNIA
18 APPETITE LOSS
19 WEIGHT LOSS
21 LIBIDO LOSS
.11
(4.9