DOCUMENT RESUME

ED 373 076 TM 021 813

AUTHOR Byrne, Barbara M.
TITLE A Comparison of EQS and LISREL Strategies in Testing for an Invariant 2nd-Order Factor Structure.
PUB DATE Apr 94
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adolescents; Comparative Analysis; Computer Oriented Programs; Computer Software; *Data Analysis; *Factor Structure; Goodness of Fit; Models; *Research Methodology; Scores; Statistical Studies
IDENTIFIERS Beck Depression Inventory; *EQS Computer Program; Invariance; LISREL Computer Program

ABSTRACT

Working from a common data base and hypothesized model, this paper demonstrates and compares the EQS and LISREL computer program strategies in the analysis of a second-order factor model. Program similarities and differences are noted with respect to: (1) preliminary analyses of the data; (2) treatment of data that are not multivariately normal; (3) assessment of overall model fit; (4) identification of parameter misspecification; (5) post hoc model-fitting; and (6) tests for multigroup invariance. Data comprise scores on the Beck Depression Inventory for 658 (337 males; 321 females) nonclinical adolescents. Issues addressed should be of substantial interest to those unfamiliar with the two problems and/or the methodological procedures presented. Includes one figure and three tables. (Contains 37 references.) (Author/SLD)

A Comparison of EQS and LISREL Strategies in Testing for an Invariant 2nd-order Factor Structure

Barbara M. Byrne

University of Ottawa

Paper presented at the American Educational Research Association Annual Meeting, New Orleans, 1994.

Abstract

Working from a common data base and hypothesized model, this paper demonstrates and compares the EQS and LISREL strategies in the analysis of a second-order factor model. Program similarities and differences are noted with respect to: (a) preliminary analyses of the data, (b) treatment of data that are not multivariately normal, (c) assessment of overall model fit, (d) identification of parameter misspecification, (e) post hoc model-fitting, and (f) tests for multigroup invariance. Data comprise scores on the Beck Depression Inventory for 658 (males, *n*=337; females, *n*=321) nonclinical adolescents. Issues addressed should be of substantial interest to those unfamiliar with the two programs and/or the methodological procedures presented.

A Comparison of EQS and LISREL Strategies in Testing for an Invariant 2nd-order Factor Structure

This past decade has seen rapid growth in the application of structural equation modeling (SEM) to data representing a wide array of disciplines. (For reviews of applications and papers related to medical and marketing research, for example, see Bentler and Stein [1992] and Bagozzi [1991], respectively.) Keeping pace with this research activity, however, has been the ongoing development and improvement of related statistical software packages. Although there are now several computer programs designed for the analysis of SEM, (e.g., CALIS [1991]; COSAN, [McDonald, 1978]; EZPATH, [Steiger, 1989]; LISCOMP [Muthen, 1988], two stand apart from the rest in terms of their popularity and widespread use. I refer, of course, to the EQS (Bentler, 1992a) and LISREL (Joreskog & Sorbom, 1993a, 1993b) programs.

Although EQS and LISREL both address the same issues related to SEM, they do so in sometimes subtle, albeit sometimes blatantly different ways. The purpose of this paper is to demonstrate a few of the dual approaches to the analysis of covariance structures as they relate to the same model and based on the same data. More specifically, using both the EQS (Version 4) and LISREL 8 (including PRELIS 2) programs, I illustrate how to (a) test for the validity of a 2nd-order factor analytic model separately for each of two groups, (b) given findings of inadequate fit, conduct post hoc model-fitting to pinpoint sources of misfit, followed by respecification and reestimation of the model, and (c) test for its invariance across the groups. Additionally, given the known kurtotic nature of the present data, I also describe the two conceptually different approaches taken by EQS and LISREL in addressing such nonnormality. Since space limitations necessarily preclude elaboration of basic principles and procedures associated with both SEM and the two statistical packages, readers are referred to Byrne (1994, 1989) for a nonmathematical approach to understanding these processes.

The Data

Data to be used in this paper are adapted from a study by Byrne, Baron, and Campbell (1993), and comprise scores on the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) for 730 adolescents (grades 9-12) attending the same high school in Ottawa Canada. Listwise deletion of data that were missing completely at random (Muthen, Kaplan, & Hollis, 1987) resulted in a final sample size of 658 (males, *n*=337; females, *n*=321).

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). For each 4-point Likert-scaled item, respondents select the statement that most accurately describes their own feelings; higher scores represent a more severe level of reported depression.

The study providing the basis for our work here is one of a series conducted by Byrne and Baron (1993a, 1993b; Byrne, Baron,

& Campbell, 1993, in press; Byrne, Baron, Larsson, & Melin, 1993a, 1993b) in validating a higher-order factorial structure of the BDI for nonclinical adolescents. Their research has demonstrated strong support for a 2nd-order structure consisting of one higher-order general factor of depression, and three lower-order factors which they labelled Negative Attitude, Performance Difficulty, and Somatic Elements. In the present paper, we examine this structure as it relates to males and females. We turn now to a more detailed view of the model under study.

The Hypothesized Model

The postulated model of BDI factorial structure is portrayed in Figure 1 in terms of both EQS and LISREL notation. It represents a typical covariance structure model and can therefore be decomposed into two submodels --a **structural model**, and a **measurement model**. The **structural model** defines the pattern of relations among the unobserved factors and is typically identified in schematic diagrams by the presence of interrelated circles, each of which represents an hypothetical construct (or factor). Turning to Figure 1, we see an hierarchical ordering of circles such that if the page were turned sideways, the "Depression" circle would be on top, with the three smaller circles beneath it. Let's now review this diagram in terms of both EQS and LISREL lexicon.

Insert Figure 1 about here

Figure 1 can be interpreted as representing one 2nd-order factor (Depression: F4; ξ_1), and three 1st-order factors (Negative Attitude: F1; η_1; Performance Difficulty: F2; η_2; Somatic Elements: F3; η_3). The single-headed arrows leading from the higher-order factor to each of the lower-order factors (F1,F4-F3,F4; γ_1-γ_3) are regression paths that indicate the prediction of Negative Attitude, Performance Difficulty, and Somatic Elements from a global Depression factor; they represent the 2nd-order factor loadings. Finally, the angled arrow leading to each 1st-order factor (D1-D3; ζ_1-ζ_3) represents residual error in the prediction of the Negative Attitude, Performance Difficulty, and Somatic Elements factors from the higher-order factor of Depression.

The **measurement model** defines relations between observed variables and unobserved hypothetical constructs. In other words, it provides the link between item scores on an assessment instrument and the underlying factors they were designed to measure. The **measurement model**, then, specifies the pattern by which each item loads onto a particular factor. This submodel can be identified by the presence of rectangular boxes, each of which represents an observed score. Turning to Figure 1 again, we see that each box represents an observed score for one BDI item. The single-headed arrows leading from each 1st-order factor to the boxes (V1-V21; λ_{11}-$\lambda_{21,3}$) are regression paths that link each of the factors to their respective set of observed scores; these coefficients (V,F's; λ's) represent the 1st-order factor loadings. For example, Figure 1 postulates that Items 16, 18, 19, and 21 load onto the Somatic Elements factor. Finally, the single-headed arrow pointing to each box (E1-E21; ϵ_1-ϵ_{21}) represents observed measurement error associated with the item variables.

One important omission in Figure 1 is the presence of double-headed arrows (↕'s) among the 1st-order factors thereby indicating their intercorrelation. This is because in 2nd-order factor analysis, all covariation among the 1st-order factors is explained by the 2nd-order factor.

Expressed more formally, the CFA model portrayed in Figure 1 hypothesized a priori that: (a) responses to the BDI could be explained by three 1st-order factors, and one 2nd-order factor of General Depression, (b) each item would have a non-zero loading on the 1st-order factor it was designed to measure, and zero loadings on the other two 1st-order factors, (c) error terms associated with each item would be uncorrelated, and (d) covariation among the three 1st-order factors would be explained fully by their regression onto the 2nd-order factor.

Assessment of Model Fit

The focal point in analyzing SEMs is the extent to which the hypothesized model "fits" or, in other words, adequately describes the sample data. This assessment entails a number of criteria, some of which bear on the fit of the model as a whole, and others, on the fit of individual parameters. Traditionally, overall model fit has been based on the χ^2 statistic. However, given the known sensitivity of χ^2 to variations of sample size, numerous alternative indices of fit have been proposed and evaluated (for reviews, see Gerbing & Anderson, 1993; Marsh, Balla, & McDonald, 1988; Tanaka, 1993). Certain of these criteria, commonly referred to as "subjective", "practical", or "ad hoc" indices of fit, are now commonly reported as adjuncts to the χ^2 statistic. We turn now to a review of these as they relate to each of the two programs. (Although both programs yield statistics related to the residual matrix, these are not included here.)

EQS Analyses

EQS reports several goodness-of-fit indices that address statistical and practical fit, as well as model parsimony. First, it yields a χ^2 statistic for both the hypothesized and independent models; the latter argues for complete independence of all variables (in this case, items) in the model. EQS also provides an optional statistic called the Satorra-Bentler χ^2 statistic (**S-B**χ^2; Satorra & Bentler, 1988). This statistic incorporates a scaling correction for the χ^2 statistic when distributional assumptions are violated.

Practical indices of fit include the Normed and Nonnormed indices (**NFI, NNFI**; Bentler & Bonett, 1980), and the Comparative fit index (**CFI**; Bentler, 1990), a revised version of the NFI that overcomes the underestimation of fit in small samples (i.e., given a correct model and small sample, the NFI may not reach 1.0 [Bentler, 1992a]). Although these three indices of fit are reported in the EQS output, Bentler (1992b) recommends the CFI to be the index of choice. Values for both the NFI and CFI range from zero to 1.00 and are derived from the comparison of an hypothesized model with the independence model; each provides a measure of complete covariation in the data, with a value >.90 indicating an acceptable fit to the data. The NNFI was originally designed to improve the NFI's performance near 1.0. However, because NNFI values can extend beyond the 0-1 range, evaluation of fit is not as readily discernible as it is with the standardized indices.

Finally, to address concerns of parsimony related to model fit, EQS provides for the evaluation of both the independent and hypothesized models based on Akaike's (1987) Information Criterion (**AIC**) and Bozdogan's (1987) consistent version of the AIC (**CAIC**); these criteria take goodness-of-fit, as well as number of estimated parameters into account.

LISREL Analyses

Versions of the program up to and including LISREL 7 included as standard output, three indices of model fit - the χ^2 statistic for the hypothesized model, the Goodness-of-fit Index (GFI), an index of the relative amount of variance and covariance jointly explained by the model, and the Adjusted GFI (AGFI) which takes into account the number of degrees of freedom in the model. In the most recent version (LISREL 8), however, the amount of model-fit information provided in the standard output has been increased dramatically to include all the goodness-of-fit measures that have been addressed in the literature (Joreskog & Sorbom, 1993a); in total, 32 evaluation criteria are reported.

In this paper, assessment of model fit for the EQS example, as it relates to single-sample analyses, is based on the S-Bχ^2 statistic and CFI*, an analog of the CFI that is computed from

S-Bχ^2 instead of χ^2 values (the S-Bχ^2 is not yet available for multigroup analyses); the LISREL example is based on the χ^2 statistic and the CFI.

Preliminary Analyses

These analyses are an essential prerequisite to SEM for several reasons. First, it is important to know if there are missing data and if so, the reason for their missingness. Given a sufficiently large sample size, and data that are missing completely at random (Muthen, Kaplan, & Hollis, 1987), listwise deletion is usually recommended when working with SEM. Second, one critically important assumption of SEM is that the data are multivariately normal. To the extent that they are not, bears on the validity of findings. While it is unlikely that the maximum likelihood estimates would be affected, nonnormality could lead to downwardly biased standard errors which would result in an inflated number of statistically significant parameters (Muthén & Kaplan, 1985). Finally, cases exhibiting extreme values of multivariate kurtosis can serve to deteriorate model fit. It is therefore important to identify and delete these outliers from the analyses.

Let's now examine sample statistics related to the present data; as noted earlier, the data are complete for both sexes.

1. Examination of Sample Statistics

EQS Analyses

When raw score data are used as input, EQS automatically provides univariate as well as several multivariate sample statistics; further insight can be obtained through descriptive analyses and the many graphical features now available in the new Windows version (Bentler & Wu, 1993) of the program. The **univariate** statistics represent the mean, standard deviation, skewness and kurtosis. As expected from previous work in this area (Byrne & Baron, 1993a, 1993b; Byrne, Baron, & Campbell, 1993, in press; Byrne et al., 1993a, 1993b), several BDI items were found to be severely kurtotic; values ranged from 0.19 to 39.40 ($\underline{M}$=4.93) for males, and from 0.15 to 10.43 ($\underline{M}$=1.92) for females.

The **multivariate** statistics reported by EQS represent variants of Mardia's (1970) coefficients of multivariate kurtosis; two reported values bear on normal theory, and two on elliptical theory. For adolescent males, the normalized estimate of Mardia's coefficient was 68.51, while for adolescent females, it was 39.49; both are distributed in very large samples from a multivariate normal population as a normal variate so that large positive values, as shown here, indicate significance.

At this time, EQS is unique in its ability to identify multivariate outliers. The program automatically prints out the five cases contributing maximally to Mardia's multivariate kurtosis coefficient. Identification of an outlier is based on the estimate presented for one case relative to those for the other four cases; there is no absolute value upon which to make this judgement, and it is possible that none of the five cases is actually an outlier; this was the case here for both adolescent males and females.

LISREL Analyses

reliminary analyses for LISREL are performed via its companion package, PRELIS. As with EQS, the input of raw data that represent continuous variables allows for the reporting of univariate statistics representing the mean, standard deviation, skewness, and kurtosis. The standard output for ordinal variables, of course, differs substantially from the one for continuous variables. While the present data are technically of ordinal measurment, they are treated as if they were continuous for purposes of consistency with the EQS analyses, as well as those of the original study. (Although EQS/Windows provides for the analysis of categorical variables, the current version of the program requires a limit of 20 variables.)

In addition to reporting the minimum and maximum frequency values, (information that is also presented in bar chart form,, PRELIS 2 also provides for single tests of zero skewness and kurtosis, as well as for an omnibus test of these two moments in combination; the single skewness and kurtosis tests are reported as z-statistics, and the omnibus test as a χ^2 statistic.

For all continuous variables jointly, PRELIS 2 similarly tests for multivariate normality. (For an extensive discussion of these tests, see Bollen, 1989). Tests for multivariate normality related to the present data revealed the following statistics for skewness (males, z=84.56; females, z=61.36), kurtosis (males, z=35.99; females, z=25.42), and for 3rd and 4th moments considered jointly (males, $\chi^2_{(2)}$=8445.74; females, $\chi^2_{(2)}$=4410.81).

2. Treatment of Nonnormality

An important assumption underlying SEM is that the data are multivariately normal. Violation of this assumption can seriously invalidate statistical hypothesis-testing such that the normal theory test statistic may not reflect an adequate evaluation of the model under study (Browne, 1982, 1984; Hu, Bentler, & Kano, 1992). One approach to resolution of the problem has been the development and use of asymptotic (large-sample) distribution-free (ADF) methods for which normality assumptions are not required (Browne, 1982, 1984). (For an extensive discussion of other solutions to the problem, see Bollen, 1989.) This is the approach embraceu by LISREL in dealing with data that are nonnormal. The strategy involves a two-step process. First, using PRELIS, the researcher recasts the data into asymptotic matrix form. LISREL analyses are then based on this matrix using weighted least squares (WLS) estimation. Nonetheless, Joreskog and Sorbom (1988a) note that the question of whether or not this approach is superior to one that uses maximum likelihood (ML) or general least squares (GLS) estimation, is still open to conjecture; furthermore, the question of how nonnormal the data must be before this process is implemented has not yet been resolved.

One major limitation associated with this treatment of nonnormality has been its excessively demanding sample size requirement. As a consequence of a major change in the storage and computation of asymptotic covariance matrices using PRELIS 2, however, the sample size restriction is now somewhat less stringent. Nevertheless, users are still cautioned that the minimum sample sizes specified by the program (for a covariance matrix, k(k+1)/2, where k = the number of variables) offer no guarantee of good estimates of the asymptotic coavriance matrix (Joreskog & Sorbom, 1993b).

Recently, however, Bentler and associates (Chou, Bentler, & Satorra, 1991; Hu et al., 1992) argued that it may be more appropriate to correct the test statistic, rather than use a different mode of estimation. As such, Satorra and Bentler (1988a, 1988b) developed the S-B χ^2 statistic which incorporates a scaling correction for the χ^2 statistic when distributional assumptions are violated; its computation takes into account the model, the estimation method, and the sample kurtosis values. From a Monte Carlo study of six test statistics under seven distributional conditions, Hu et al., reported the S-Bχ^2 to be the most reliable. This is the approach taken by the EQS program in the treatment of nonnormal data. In contrast to LISREL, then, EQS uses an estimation method that assumes the data are multivariate normal, but bases evaluation of model fit on a test statistic that has been corrected to take nonnormality into account.

Testing the Hypothesized Model of BDI Structure

A summary of selected fit indices for both the EQS and LISREL analyses is presented in Table 1. Results are reported both for analyses that took the nonnormality of the data into account, and for those based on normal theory estimation (i.e., data were considered to be normally distributed). ML estimation was used for all analyses except those based on nonnormal data using LISREL 8; the latter were based on ADF estimation as recommended by Joreskog and Sorbom (1988a). Not unexpectantly (see Hu et al., 1992; Joreskog & Sorbom, 1988a), the LISREL model-fitting results based on ADF estimation are somewhat at odds with the findings based on ML estimation. Although the basic pattern is similar, the χ^2 (as a measure of bad fit) and CFI (as a measure of good fit) values are excessively high. One possible explanation of the latter may lie with the enormous χ^2 value for the highly misspecified null model; this of course, would lead to an inflated CFI value. Interpretation of findings, then, are therefore limited to the ML estimates and are based on the S-Bχ^2 and CFI* for EQS, and on the χ^2 and CFI for LISREL.

Insert Table 1 about here

As indicated by the CFI* (EQS), and CFI (LISREL) values reported in Table 1, goodness-of-fit for the initially hypothesized model of BDI structure was exceptionally good for males; it was somewhat less so for females. However, before turning to the problematic fit for adolescent females, let's first complete our evaluation of the hypothesized model for adolescent males by assessing the fit of individual parameters in the model. For both EQS and LISREL, there are two aspects of concern here: (a) the appropriateness of the estimates, and (b) their statistical significance. Any differences between the two programs are noted below in the discussion of these criteria.

Feasibility of Parameter Estimates. The first step in assessing the fit of individual parameters is to determine the viability of their estimated values. Any estimates falling outside the admissable range signal that either the model is wrong, or the input matrix lacks sufficient information. Examples of parameters exhibiting unreasonable estimates are: (a) correlations >1.00, (b) standard errors that are abnormally large or small. A standard error approaching zero usually results from the linear dependence of the related parameter, with some other parameter in the model; such a circumstance renders testing for the statistical significance of the estimate impossible, and (c) negative variances. Whereas LISREL permits these estimates to be printed, EQS prevents their estimation by constraining the value of the offending parameter to zero; the message **PARAMETER XX,XX CONSTRAINED AT LOWER BOUND** will appear on the output.

Statistical Significance of Parameter Estimates. The test statistic here represents the parameter estimate divided by its standard error; as such, it operates as a z-statistic in testing that the estimate is statistically different from zero. Based on an α level of .05, then, the test statistic needs to be > ±1.96 before the hypothesis (that the estimate=0.0) can be rejected. LISREL 7 and its predecessors referred to these values as "t-values". The output for LISREL 8, however, is consistent with that of EQS in reporting these test statistics, and their standard errors, immediately under each parameter estimate. One additional difference between the two programs is that if the EQS user requested robust statistics (i.e., S-Bχ^2), the output will report two sets of test statistics and standard errors - one for the original, and one for the corrected χ^2 statistics.

For purposes of comparison across programs and estimation processes, EQS and LISREL estimates are presented in Table 2. In consideration of space, however, only the 1st-order factor loading estimates are reported, and as they pertain only to adolescent males. With respect to the previous point, note that while the maximum likelihood estimate (under normal theory) for Item 19 was significant, it was not so when multivariate kurtosis was taken into account by the S-Bχ^2 statistic reported by the EQS program.

Insert Table 2 about here

Post Hoc Model-fitting to Establish Baseline Models

When an hypothesized model is tested and the fit found to be inadequate, it is customary to proceed with post hoc model-fitting to identify misspecified parameters in the model. If multigroup equivalence is of interest, it is particularly important that a baseline model be established for each group separately before testing for their invariance across groups. This model represents one that is most parsimonious, as well as statistically best-fitting and substantively most meaningful. Identification of misspecified parameters differs substantially between the EQS and LISREL programs. Whereas EQS takes a multivariate approach based on the **Lagrange Multiplier Test (LM-Test)**, the LISREL approach is univariate and is based upon the

Modification Index (MI). Nonetheless, the objective of both tests is to determine if a model that better represents the data would result with certain parameters specified in subsequent run as free, rather than fixed.

Before putting these techniques into practice, however, one vitally important caveat needs to be stressed with respect to use of both the LM-Test and MIs in the respecification of models. It bears on two factors: (a) that both techniques are based solely on statistical criteria, and (b) that virtually any fixed parameter (constrained either to zero, or some nonzero value) is eligible for testing. Thus, it is critical that the researcher pay close heed to the substantive theory before relaxing constraints as may be suggested by both the LM and MI statistics; model respecification in which certain parameters have been set free **must** be substantiated by sound theoretical rationale!

Let's now return to the problematic fit of BDI structure for adolescent females and examine these differential posthoc model-fitting procedures within the context of the two statistical packages.

EQS Analyses

Examination of the multivariate LM χ^2 coefficients related to the initially hypothesized model (Model 1) for females revealed substantial improvement in model fit to be gained from the additional specification of an error covariance between Items 21 and 20 ($LM\chi^2_{(1)}$=22.59), and the cross-loading (the loading of a single item on more than one factor) of Item 20 on the higher-order factor of Depression ($LM\chi^2_{(2)}$=15.81) (i.e., Item 20 loaded on F_4 as well as on F_2).

Since the loading of Item 20 onto the Depression factor would lead to a psychometrically ambiguous specification, the model was reparameterized as a 1st-order CFA model in order to assess possible misspecification at the lower structural level. Estimation of this model replicated the misspecification of both the error covariance and Item 20; the latter was shown to cross-load on Factor 1. Thus, the hypothesized model (Model 1) for females was respecified to include these two additional parameters, and then reestimated. That we were able to reparameterize the model by respecifying multiple parameters in a single run represents a major difference from the LISREL program, where only one parameter can be respecified at a time. As a consequence, this respecified model represents Model 3 in Table 1, since Model 2 is redundant to the EQS analyses.

To assess the extent to which each newly specified model exhibits an improvement over its predecessor, we examine the difference in χ^2 ($\Delta\chi^2$) between the two nested models. This differential i- itself χ^2-distributed, with degrees of freedom equal to the difference in degrees of freedom (Δdf) and can, thus, be tested statistically; a significant $\Delta\chi^2$ indicates a substantial improvement in model fit. As is evident in Table 1, the inclusion of these two parameters in the model yielded a statistically significant and substantial improvement in model fit (ΔS-B$\chi^2_{(2)}$=31.33; ΔCFI*=.04). Closer scrutiny of the parameter estimates, however, revealed the original loading of Item 20 on Factor 2 to be nonsignificant. In the interest of parsimony, then, the model was respecified with this parameter deleted. Because Model 4 was deemed to be substantively reasonable (see Byrne et al., 1993 for an extended explanation) and exhibited an excellent fit to the data, it was considered the most plausible in representing the data for adolescent females.

LISREL Analyses

Consistent with the EQS analyses, the LISREL results based on ML estimation also yielded a better-fitting model for males, than for females, as indicated by CFI value <.90 reported in Table 1. A review of the MIs revealed two parameters to be potentially worthy of estimation. The more prominent fixed parameter (MI=22.60) represented the error covariance between Items 21 and 20; the other (MI=17.04) represented the cross-loading of Item 20 onto the Negative Attitude factor. As shown in Table 1, three separate models (Models 2-4) were subsequently specified and estimated.

A review of results related to these models reveals each to yield a highly significant improvement in model fit over its predecessor. As with the EQS analyses, for statistical, psychometric, and theoretical reasons, Model 4 was considered to be the most plausible in representing BDI data for adolescent females.

Testing for Invariance Across Gender

Having determined the baseline model for each sex, analyses proceeded next to test for their factorial equivalence across males and females. At first blush, except for the differential loading pattern of Item 20 and the specification of an error covariance for females, one might be quick to conclude that the BDI was factorially equivalent across gender. Such a conclusion would be premature, however, since a similarly specified model in no way guarantees the equivalence of item measurements and underlying theoretical structure; related hypotheses must be tested statistically in a simultaneous analysis of data from both groups. We turn now to these analyses as they are addressed separately within the EQS and LISREL programs.

EQS Analyses

Since we already know prior to testing for cross-group invariance, that Item 20 is apparently perceived differently by adolescent males and females, the factor loading for this item was not constrained equal across gender; the error covariance is also unique to females, and is free to take on any value. Such specification addresses the issue of partial measurement invariance in the testing of equivalence across multiple samples (see Byrne, Shavelson, & Muthen, 1989).

In EQS, we can test for the invariance of both the 1st- and 2nd-order factor loadings simultaneously. This approach is made possible in two important ways. First, it employs the multivariate LM-Test in the evaluation of equality constraints, and second, it makes the detection of misspecified constraints easy by providing probability values associated with the LM χ^2 statistic for each. A review of these statistics revealed four constraints to be untenable. Probability values <.05 were associated with Items 8, 10, 12, and 18 thereby arguing for their nonequivalence across adolescent males and females.

LISREL Analyses

Testing for invariance based on LISREL involved the testing of three increasingly restrictive hypotheses, each nested within the one preceding; these related to the equivalency of (a) number of underlying factors, (b) 1st-order factor loadings, and (c) 2nd-order factor loadings. (For an elaboration of this procedure, see Byrne, 1989.)

Analyses involved specifying a model in which certain parameters were constrained equal across gender, and then comparing that model with a less restrictive one in which the same parameters were free to take on any value. As with model-fitting, the $\Delta\chi^2$ between competing models provided a basis for determining the tenability of the hypothesized equality constraints; a significant $\Delta\chi^2$ indicating noninvariance (i.e., nonequivalence). Turning to the summary of LISREL analyses shown in Table 3, we see that the first invariance model (Model 1) tested for the equivalence of an underlying 3-factor structure (irrespective of factor loading pattern) across males and females. This initial specification simply tests for adequacy of model fit in a simultaneous analysis of multigroup data, and provides the criterion against which the two subsequent invariance models are compared; given a CFI value of .92, multigroup model fit was considered to be reasonably good. A second model was then specified in which the pattern of lower-order factor loadings was constrained equal across the two groups. (Note that Item 20 was not constrained equal across groups). Comparison of this model (Model 2) with Model 1 yielded a statistically significant difference in model fit ($p<.01$), thereby substantiating rejection of the hypothesis that item measurements were equivalent across males and females.

Insert Table 3 about here

Given findings of some gender specificity related to the lower-order factors, the next task was to identify the BDI items contributing to this noninvariance. This was accomplished by first testing separately for the invariance of each BDI subscale (i.e., all items comprising each subscale were tested as a group). Given significant findings for any one of these three tests, analyses proceeded next in testing for the invariance of each item within each subscale. Finally, constraining all 1st-order loadings known to be group-invariant, analyses then focus on the 2nd-order factor loadings. Due to limitations of space, results related to these nested series of tests are simply summarized, as shown in Table 3. Readers who may wish a more detailed description of this model-testing procedure are referred to Byrne (1989, 1994; Byrne et al., 1989).

Summary

Working from a common data base and hypothesized model, this paper has provided an extant example of the EQS and LISREL strategies in testing for an invariant 2nd-order factor structure across groups. Along the way, similarities and differences between the two programs were noted with respect to: (a) approach to, and information derived from preliminary analyses of the data, (b) treatment of data that violate the assumption of multivariate normality, (c) assessment of overall model fit, (d) identification of parameter misspecification, (e) post hoc model fitting, and (f) tests for multigroup invariance.

Although, substantively, results based on ML estimation were consistent across the two programs, those bearing on the equality of BDI measurement and structure across groups differed with respect to five parameters - four 1st- and one 2nd-order loadings. The discrepancy in these findings is undoubtedly a consequence of the univariate versus multivariate approach to the identification of misspecified equality constraints taken by LISREL and EQS, respectively. Of most concern is the inconsistent finding related to the 2nd-order loading of F_3 on F_4. One explanation likely lies in the highly correlated structure among the 1st-order factors for both males (mean r=.78) and females (mean r=.76) which would not be taken into account in the univariate test for invariance.

EQS and LISREL model fit statistics related to analyses that took the nonnormality of the data into account were widely0 discrepant. Whereas the EQS approach in correcting the χ^2 statistic yielded results that were reasonable, the χ^2 statistic and CFI value produced by LISREL, based on the ADF estimator, were unreasonably high. These findings support those from a Monte Carlo study reported by Hu and colleagues (1992) that revealed the ADF statistic to perform as a χ^2 variate only when sample size approximates 5,000 cases. Given that most practical applications of SEM involve substantially smaller sample sizes, the S-Bχ^2 statistic produced by EQS appears to be the more useful measure of model fit when the data are in violation of the normality assumption.

Although this comparison of the EQS and LISREL programs has highlighted only a few of their differential approaches to SEM application, it is hoped that the issues addressed here will be helpful to readers who may be relatively unfamiliar with the two programs and/or the methodological procedures presented.

References

Bagozzi, R.P. (1991). Structural equation modeling in marketing research. In W.D. Neal (Ed.), First annual advanced research techniques forum (pp. 335-379), Chicago: American Marketing Association.

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.

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.

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

Bentler, P.M. (1992a). EQS: Structural Equations program manual. Los Angeles, CA: BMDP Statistical Software.

Bentler, P.M. (1992b). On the fit of models to covariances and methodology to the **Bulletin**. Psychological Bulletin, 112, 400-404.

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. & Stein, J.A. (1992). Structural equation modeling in medical research. Statistical Methods in Medical Research, 1, 159-181.

Bentler, P.M. & Wu, E.J.C. (1993). EQS/Windows user's guide. Los Angeles, CA: BMDP Statistical Software.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (Ed.), Topics in multivariate analysis (pp. 72-141). Cambridge, Eng.: Cambridge University Press.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematics and Statistical Psychology, 37, 62-83.

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

Byrne, B.M. (1994). Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming. Newbury Park, CA: Sage.

Byrne, B.M. & Baron, P. (1993a). Measuring adolescent depression: Tests of equivalent factorial structure for English and French versions of the Beck Depression Inventory. Applied Psychology: An International Review, 43, 33-47.

Byrne, B.M. & Baron, P. (1993b). The Beck Depression Inventory: Testing and cross-validating an hierarchical factor structure for nonclinical adolescents. Measurement and Evaluation in Counseling and Development, 26, 164-178.

Byrne, B.M., Baron, P., & Campbell, T.L. (1993). Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender. Journal of Research on Adolescence, 3, 127-143.

Byrne, B.M., Baron, P., & Campbell, T.L. (in press). The Beck Depression Inventory (French version): Testing for gender-invariant factorial structure for nonclinical adolescents. Journal of Adolescent Research.

Byrne, B.M., Baron, P., Larsson, B., & Melin, L. (1993a). The Beck Depression Inventory: Testing and cross-validating a second-order structure for Swedish nonclinical adolescents. Under review.

Byrne, B.M., Baron, P., Larsson, B., & Melin, L. (1993b). Measuring depression for Swedish nonclinical adolescents: Factorial validity and equivalence of the Beck Depression Inventory across gender. Under review.

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

Chou, C-P, Bentler, P.M., & Satorra, A (1991). Scaled test statistics and robust standard errors for nonnormal data in covariance structure analysis: A Monte Carlo study. British Journal of Mathematical and Statistical Psychology, 44, 347-357.

Gerbing, D.W. & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen & J.S. Long (Eds.) Testing structural equation models. Newbury Park, CA: Sage.

Hu, L.-, Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? Psychological Bulletin, 112, 351-362.

Joreskog, K.G. & Sorbom, D. (1988a). LISREL 7: A guide to the program and applications. Chicago, IL: SPSS Inc.

Joreskog, K.G. & Sorbom, D. (1988b). PRELIS: A preprocessor for LISREL. Mooresville, IN: Scientific Software Inc.

Joreskog, K.G. & Sorbom, D. (1993a). LISREL 8: User's reference guide. Chicago, IL: Scientific Software International Inc.

Joreskog, K.G. & Sorbom, D. (1993b). PRELIS 2: User's reference guide. Chicago, IL: Scientific Software International Inc.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530.

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.

McDonald, R.P. (1978). A simple comprehension model for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 31, 59-72.

Muthen, B. (1988). LISCOMP: Analysis linear structural equations with a comprehensive measurement model. Chicago: Scientific Software.

Muthen, 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.

Muthen, B., Kaplan, D., & Hollis, M (1987). Structural equation modeling for data that are not missing completely at random. Psychometrika, 51, 431-462.

SAS Institute Inc. (1991). The CALIS procedure: Analysis of covariance structures. Cary, NC: Author.

Satorra, A. & Bentler, P.M. (1988). Scaling corrections for chi-square in covariance structure analysis. Proceedings of the American Statistical Association.

Steiger, J.H. (1989). EZPATH: A supplementary module for SYSTAT and SYGRAPH. Evanston. IL: SYSTAT.

Tanaka, J.S. Multifaceted conceptions of fit in structural equation models. In K.A. Bollen & J.S. Long (Eds.), Testing structural equation models. Newbury Park, CA: Sage.

Table 1

Summary of Goodness-of-fit Statistics for Adolescent Males and Females

		Nonnormality Taken into Account					Nonnormality Not Taken into Account			
		EQS[a]			LISREL[b]		EQS[a]		LISREL[a]	
Model		df	S-B χ^2	CFI*	χ^2	CFI	χ^2	CFI	χ^2	CFI
Males										
0	Null[c]	210	906.13	—	NA	—	NA	—	NA	—
1	Hypothesized	187[d]	224.62	.95	465.66	.98	306.50	.92	306.49	.91
Females										
0	Null	210	1136.75	—	NA	—	NA	—	NA	—
1	Hypothesized	187[d]	266.66	.91	667.11	.95	340.16	.88	340.16	.88
2	Model 1 with covariance between items 20 and 21	186	NA	NA	663.63	.95	NA	NA	316.80	.90
3	Model 2 with Item 20 cross-loaded on F_1	185	235.33	.95	650.57.	.95	296.07	.92	296.07	.92
4	Model 2 with Item 20 loaded on F_1 instead of F_2	186	236.52	.95	650.82	.95	297.69	.92	297.69	.92

[a]Based on maximum likelihood estimation

[b]Based on the large-sample asymptotic matrix with weighted least squares estimation

[c]Required for calculation of CFI*

[d]Disturbance terms associated with F_2 and F_3 constrained equal for purposes of statistical identification

F_1 = Factor 1 (Negative Attitude); F_2 = Factor 2 (Performance Difficulty)

NA = Not Applicable

Table 2

Summary of EQS and LISREL First-order Factor-Loading Estimates for Adolescent Males

	Nonnormality Taken into Account						Nonnormality Not Taken into Account					
	EQS[b]			LISREL[c]			EQB[b]			LISREL[b]		
BDI Items	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Item 1	1.00			1.00			1.00			1.00		
Item 2	1.08(.14)			0.90(.06)			1.08(.12)			1.08(.12)		
Item 3	1.10(.18)			0.82(.06)			1.10(.12)			1.10(.12)		
Item 5	0.87(.15)			0.76(.05)			0.87(.10)			0.87(.10)		
Item 6	1.06(.02)			1.04(.09)			1.06(.15)			1.06(.15)		
Item 7	1.16(.13)			1.00(.05)			1.16(.12)			1.16(.12)		
Item 8	1.30(.18)			1.10(.07)			1.30(.14)			1.30(.14)		
Item 9	0.79(.16)			0.72(.06)			0.79(.10)			0.79(.10)		
Item 10	0.71(.15)			0.87(.10)			0.71(.18)			0.71(.18)		
Item 14	0.90(.18)			0.72(.07)			0.90(.12)			0.90(.12)		
Item 4		1.00			1.00			1.00			1.00	
Item 11		0.61(.14)			1.07(.15)			0.61(.14)			0.61(.14)	
Item 12		0.67(.12)			0.87(.08)			0.67(.09)			0.67(.09)	
Item 13		1.01(.12)			1.50(.12)			1.01(.13)			1.01(.13)	
Item 15		0.82(.10)			1.10(.11)			0.82(.11)			0.82(.11)	
Item 17		0.80(.10)			0.92(.09)			0.80(.11)			0.80(.11)	
Item 20		0.67(.12)			0.80(.08)			0.67(.10)			0.67(.10)	
Item 16			1.00			1.00			1.00			1.00
Item 18			1.13(.27)			0.46(.06)			1.13(.26)			1.13(.26)
Item 19			0.28(.15)*			0.16(.03)			0.28(.12)			0.28(.12)
Item 21			0.58(.29)			0.12(.01)			0.58(.14)			0.58(.14)

* Not significant

[a] Standard errors in parentheses

[b] Based on maximum liklihood estimation

[c] Based on the large-sample asymptotic matrix with weighted least squares estimation

F_1 = Factor 1 (Negative Attidude); F_2 = Factor 2 (Performance Difficulty); F_3 = Factor 3 (Somatic Elements)

Table 3
Summary of LISREL Tests for Invariance Across Gender

Model	χ^2	df	CFI	Model Comparison	$\Delta\chi^2$	Δdf
1 Baseline Multigroup model	604.18	373	.92	---	---	---
2 All 1st-order factor loadings invariant[a]	641.25	390	.91	2 vs 1	37.07**	17
3 Item loadings for F_1 invariant	626.88	382	.91	3 vs 1	22.70**	9
4 Item loadings for F_2 invariant	610.91	378	.91	4 vs 1	6.73	5
5 Item loadings for F_3 invariant	611.87	376	.91	5 vs 1	7.69	3
6 All 1st-order factor loadings invariant[b] except Items 8 and 20	632.70	389	.91	6 vs 1	28.52*	16
7 all 1st-order factor loadings invariant except Items 8,19, 20	627.85	388	.91	7 vs 1	23.67	15
8 Model 7 with all 2nd-order loadings invariant	636.01	391	.91	8 vs. 7	8.16*	3
9 Model 7 with 2nd-order loadings for F_1 and F_2 invariant[b]	628.41	390	.91	9 vs. 7	0.56	2

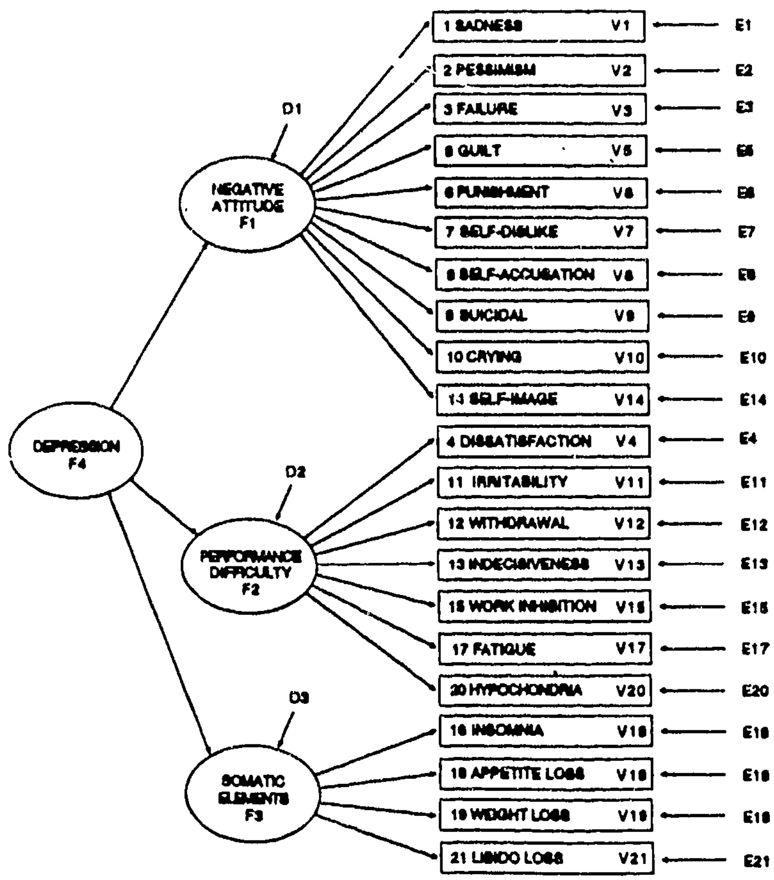
*$p<.05$ ** $p < .01$
[a] Item #20 was not constrained equal across gender.
[b] Equality constraints were imposed separately for each item loading.

$\Delta\chi^2$ = difference in χ^2 values; Δdf = difference in degrees of freedom.
F1 = Factor 1 (Negative Attitudes); F2 = Factor 2 (Performance Difficulty); F3 = Factor 3 (Somatic Elements); CFI = Comparative Fit Index

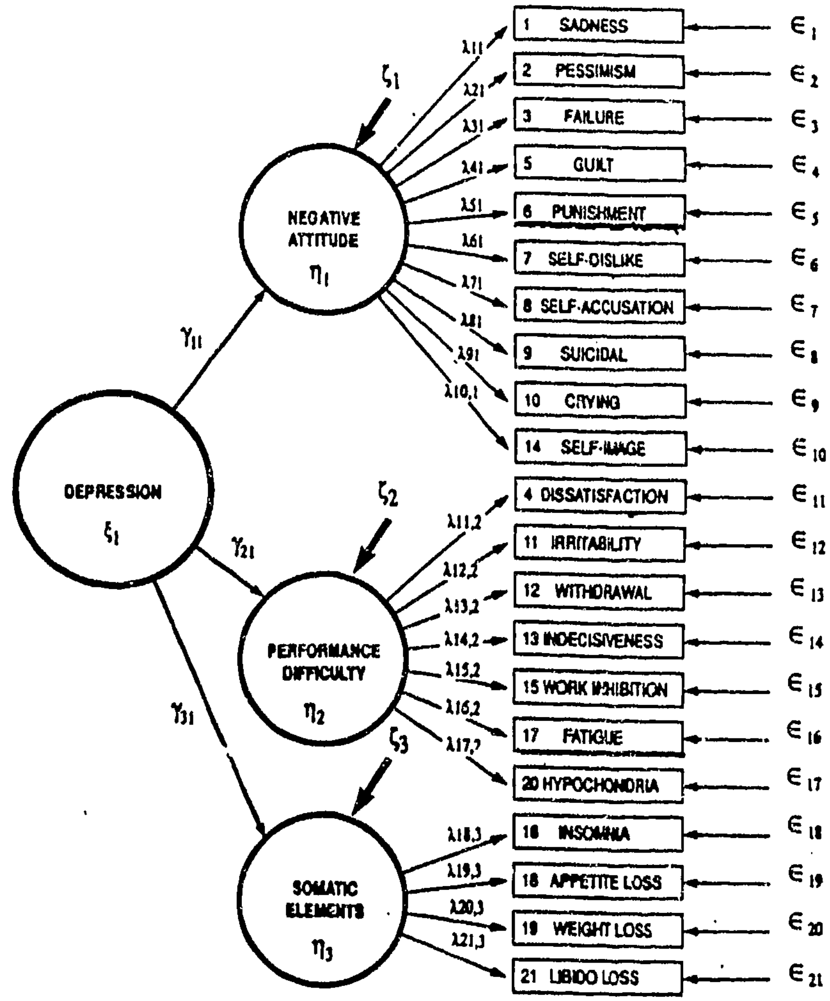
Figure Caption

<u>Figure 1</u>. Hypothesized 2nd-order Model of BDI Factorial Structure Expressed in both EQS and LISREL Notation

EQS


DEPRESSION F4
NEGATIVE ATTITUDE F1
PERFORMANCE DIFFICULTY F2
SOMATIC ELEMENTS F3
D1
D2
D3
1 SADNESS V1 E1
2 PESSIMISM V2 E2
3 FAILURE V3 E3
5 GUILT V5 E5
6 PUNISHMENT V6 E6
7 SELF-DISLIKE V7 E7
8 SELF-ACCUSATION V8 E8
9 SUICIDAL V9 E9
10 CRYING V10 E10
14 SELF-IMAGE V14 E14
4 DISSATISFACTION V4 E4
11 IRRITABILITY V11 E11
12 WITHDRAWAL V12 E12
13 INDECISIVENESS V13 E13
15 WORK INHIBITION V15 E15
17 FATIGUE V17 E17
20 HYPOCHONDRIA V20 E20
16 INSOMNIA V16 E16
18 APPETITE LOSS V18 E18
19 WEIGHT LOSS V19 E19
21 LIBIDO LOSS V21 E21

LISREL


DEPRESSION ξ_1
NEGATIVE ATTITUDE η_1
PERFORMANCE DIFFICULTY η_2
SOMATIC ELEMENTS η_3
γ_{11}
γ_{21}
γ_{31}
ζ_1
ζ_2
ζ_3
λ_{11} 1 SADNESS ϵ_1
λ_{21} 2 PESSIMISM ϵ_2
λ_{31} 3 FAILURE ϵ_3
λ_{41} 5 GUILT ϵ_4
λ_{51} 6 PUNISHMENT ϵ_5
λ_{61} 7 SELF-DISLIKE ϵ_6
λ_{71} 8 SELF-ACCUSATION ϵ_7
λ_{81} 9 SUICIDAL ϵ_8
λ_{91} 10 CRYING ϵ_9
$\lambda_{10,1}$ 14 SELF-IMAGE ϵ_{10}
$\lambda_{11,2}$ 4 DISSATISFACTION ϵ_{11}
$\lambda_{12,2}$ 11 IRRITABILITY ϵ_{12}
$\lambda_{13,2}$ 12 WITHDRAWAL ϵ_{13}
$\lambda_{14,2}$ 13 INDECISIVENESS ϵ_{14}
$\lambda_{15,2}$ 15 WORK INHIBITION ϵ_{15}
$\lambda_{16,2}$ 17 FATIGUE ϵ_{16}
$\lambda_{17,2}$ 20 HYPOCHONDRIA ϵ_{17}
$\lambda_{18,3}$ 16 INSOMNIA ϵ_{18}
$\lambda_{19,3}$ 18 APPETITE LOSS ϵ_{19}
$\lambda_{20,3}$ 19 WEIGHT LOSS ϵ_{20}
$\lambda_{21,3}$ 21 LIBIDO LOSS ϵ_{21}