DOCUMENT RESUME

ED 278 689 TM 870 106

AUTHOR	Marsh, Herbert W.
TITLE	The Factorial Invariance of Responses by Males and Females to a Multidimensional Self-Concept Instrument: Substantive and Methodological Issues.
PUB DATE	1 Mar 86
NOTE	29p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Achievement; Correlation; *Factor Analysis; *Factor Structure; Foreign Countries; *Goodness of Fit; Grade 5; Intermediate Grades; Mathematical Models; Research Methodology; *Self Concept; *Self Concept Measures; Sex Differences
IDENTIFIERS	Australia; *Self Description Questionnaire

ABSTRACT

The purposes of this study are to examine the factorial invariance of responses by preadolescent males and females to a multidimensional self-concept instrument, and to demonstrate the use of confirmatory factor analysis (CFA). Sets of responses by 500 males and by 500 females were each randomly divided in half to form four groups (M1, M2, F1, and F2). The factorial invariance of an a priori structure demonstrated the replicability of the structure across the random split halves (M1 and M2, and F1 and F2), and the generality of the structures across opposite-sex comparisons (M1 and F1, and M2 and F2). Additional a posteriori structures that better fit the data were derived on the basis of the initial analyses, but the estimated values of the new parameters were not strictly invariant across either random split halves or opposite-sex comparisons. This suggests that some of the improved fit was illusory and due to capitalizing on chance. However, all opposite-sex comparisons demonstrated the invariance of factor loadings and factor correlations for a priori and a posteriori structures. A three-page list of bibliographical references and the tabulated data follow the report. (Author/JAZ)

The Factorial Invariance of Responses by Males and Females to a Multidimensional Self-concept Instrument: Substantive and Methodological Issues

Herbert W. Marsh
The University of Sydney, Australia

1 March, 1986

Running Head: Factorial Invariance

The Factorial Invariance of Responses by Males and Females to a Multidimensional Self-concept Instrument: Substantive and Methodological Issues

ABSTRACT

The purposes of this study are to examine the factorial invariance of responses by preadolescent males and females to a multidimensional self-concept instrument, and to demonstrate the use of confirmatory factor analysis (CFA). Sets of responses by 500 males and by 500 females were each randomly divided in half to form four groups (M1, M2, F1, and F2). The factorial invariance of an a priori structure demonstrated the replicability of the structure across random split halves (M1 & M2, and F1 & F2), and the generality of the structures across opposite-sex comparisons (M1 & F1, and M2 & F2). Additional a posteriori structures that better fit the data were derived on the basis of the initial analyses, but the estimated values of the new parameters were not strictly invariant across either random split halves or opposite-sex comparisons. This suggests that some of the improved fit was illusory and due to capitalizing on chance. However, all opposite-sex comparisons demonstrated the invariance of factor loadings and factor correlations for a priori and a posteriori stfuctures.

The Factorial Invariance of Responses by Males and Females to a Multidimensional Self-concept Instrument: Substantive and Methodological Issues

The Substantive and Methodological Issues of the Study.

The purposes of this study are to: a) examine the factorial invariance of responses by preadolescent males and females to a multidimensional self-concept instrument, a substantive issue; and b) demonstrate the use of confirmatory factor analysis (CFA) for tests of factorial invariance and to examine potential problems with its use and interpretation, a methodological issue.

The Structure of Self-concept: The Substantive Issue.

Self-concept research has suffered from a paucity of theoretical models and psychometrically-sound measurement instruments. Shavelson, Hubner, and Stanton (1976) reviewed theoretical and empirical research, and developed a multifaceted model of self-concept that served as the basis of the Self Description Questionnaire (SDQ) used in the present investigation. Through the mid-1970s self-concept instruments typically consisted of a hodge-podge of self-referent items, and "blind" applications of exploratory factor analysis (EFA) typically failed to identify salient, replicable factors (see Marsh & Smith, 1982; Shavelson, et al., 1976). More recently, researchers have developed instruments to measure specific self facets that are at least loosely based on explicit theoretical models such as proposed by Shavelson, and then used factor analyses to support these a priori factors (Boersma & Chapman, 1979; Dusek & Flaherty, 1981; Fleming & Courtney, 1984; Harter, 1982; Marsh, Barnes, Cairns & Tidman, 1984; Marsh & Hocevar, 1985; Marsh, Smith & Barnes, 1985; Soares & Soares, 1982). Recent reviews of such research (Byrne, 1984; Marsh & Shavelson, 1985) support the multidimensional structure of self-concept, and emphasize research based on the SDQ.

Sex differences are one of the most frequently examined influences on self-concept responses (e.g., Wylie, 1979; Marsh, Barnes, et al., 1984; Marsh, 1985a; 1986a). However, few researchers have examined the factor structures of responses by males and females, and unless the factor structures are similar then there may be no basis for comparing mean responses by males and females. Marsh, Barnes, et al. (1984) examined sex and age effects in responses to the SDQ by preadolescent students in grades 2-5 (also see Marsh, 1985a; 1986a), and illustrated a moderate degree of similarity in factor structures identified by EFA across the four age groups, but did not examine separate factor analyses for responses by males and females. Dusek & Flaherty (1981) used EFA to show that factor structures for their self-concept instrument were similar across age groups and responses by males and females. However, Marsh and Hocevar (1985) argued that EFA was not entirely appropriate for the comparison of factor structures (see discussion below) and used confirmatory factor analysis (CFA) to demonstrate that factor loadings used to define each factor was invariant across age groups. Consistent with the Shavelson model, they found that the size of correlations among the factors varied systematically with age. The substantive focus of the present investigation is to examine the invariance of responses to the SDQ across responses by males and females with CFA.

<u>Factorial Invariance -- The Methodological Issue.</u>

Historically, EFA has been used to examine the similarity of two independent factor structures. The similarity of solutions based on the same measured variables for similar groups is used to infer the replicability of the factor solution, while the similarity across dissimilar groups is an indication of the generalizability of the factor solution. A wide variety of comparison procedures based on the similarity of the factor loadings from EFAs have been proposed (see Gorsuch, 1974; Harman, 1967; Everett, 1983; Everett & Entrekin, 1980). While EFA continues to be widely used, as typically applied it imposes many undesirable limitations. First, the researcher is unable to define a particular factor structure beyond determining the number of factors to be rotated, and perhaps the obliqueness of the rotated factors. Second, the factor solution is not unique and mathematically equivalent solutions with different interpretations are plausible. When the observed factor structure does not closely correspond to the hypothesized structure, there is no way of determining the extent to which the hypothesized structure is able to fit the data. These limitations are even more serious when trying to compare factor solutions from two or more groups; the researcher is not able to define the factor structure for either group or to specify that the structure be the same across the groups. If both solutions do not closely match the hypothesized structure or each other, then there is no way to determine how well the hypothesized structure is able to fit the data from either group or how well solutions from each group match each other. Alwin and Jackson (1981) further noted that the investigation of factorial invariance with EFA confounds separate and distinct issues such as the invariance of specific aspects of the factor solution, and argued that "the use of exploratory factor analysis in its conventional form to examine issues of factorial invariance is of limited utility" (p. 253).

In CFA the researcher defines the specific factor structure to be examined, and is able to test its ability to fit the data. Recent methodological advances in the application of CFA (Alwin & Jackson, 1981; Joreskog & Sorbom, 1981; Marsh, 1985b; Marsh & Hocevar, 1985) provide a more rigorous comparison of the factor structures resulting from multiple groups where the researcher is able to test the fit of a model in which any specified group of parameter estimates are constrained to be invariant across groups. This allows the researcher to specify the factor structure to be examined, to uniquely identify parameters in the solutions, and to test hypotheses of invariance for particular components of the factor solution. Here, the researcher is not only examining the similarity of the pattern of parameter estimates from two different groups, but is testing whether the actual values of the parameters are the same across groups.

Methods.

The Self Description Questionnaire.

The SDQ assesses three areas of academic self-concept and four areas of nonacademic self-concept derived from the Shavelson model of self-concept (Shavelson, et al. 1976) as well as a General-self derived from the Rosenberg (1965) self-esteem scale. On the SDQ, preadolescent children are asked to respond to simple declarative sentences (e.g., I'm good at mathematics, I make friends easily) with one of five response categories: false; mostly false; sometimes false, sometimes true; mostly true; true. Each of the eight SDQ scales is inferred on the basis of eight positively worded items. For all students in the present investigation, the SDQ was administered to intact classes of students during regular school hours according to standardized administration procedures that are presented in the test manual (Marsh, 1986a). A brief description of the eight SDQ scales is as follows:

1) Physical Abilities/Sports (Phys) -- student ratings of their ability and enjoyment of physical activities, sports and games;

2) Physical Appearance (Appr) -- student ratings of their own attractiveness, how their appearance compares with others, and how others think they look;

3) Peer Relations (Peer) -- student ratings of how easily they make friends, their popularity, and whether others want them as a friend;

4) Parent Relations (Prnt) -- student ratings of how well they get along with their parents and whether they like their parents;

5) Reading (Read) -- student ratings of their ability in and their enjoyment/interest in reading;

6) Mathematics (Math) -- student ratings of their ability in and their

enjoyment/interest in mathematics;

7) General-school (Schl) -- student ratings of their ability in and their enjoyment/interest in "all school subjects;"

8) General-self (Genr) -- student ratings of themselves as effective, capable individuals, who are proud and satisfied with the way they are.

Descriptions of the instrument, the theoretical definition of self-concept upon which it is based, the eight scales, internal consistency estimates of reliability, numerous EFAs and CFAs of responses to the SDQ, and construct validity studies are summarized in the test manual (Marsh, 1986a; also see Marsh, 1985a; 1986b; Marsh, Barnes, et al., 1984; Marsh & Hocevar, 1985; Marsh & Parker, 1984; Marsh, Parker & Smith, 1983; Marsh, Relich & Smith, 1984; Marsh & Richards, 1986; Marsh & Shavelson, 1985; Marsh, Smith & Barnes, 1983; 1984; 1985; Marsh, Smith, Barnes & Butler, 1983). This research has shown the SDQ scales to be well defined, quite distinct (mean r = .16), reliable (coefficient alphas in the .80s and .90s), moderately correlated with measures of corresponding academic abilities (.3 to .7), in agreement with self-concepts inferred by others, affected by experimental manipulations designed to enhance self-concepts, and logically related to other constructs.

<u>Data For the Present Investigation.</u>

Data for the present investigation come from the normative archive of responses to the SDQ by preadolescents in grades 2 to 6 as described in the SDQ test manual (Marsh, 1986a). For purposes of the present investigation responses by 1000 fifth graders -- 500 males (M) and 500 females (F) -- were randomly selected, and each set of 500 responses was then randomly divided into half to form four groups of 250 called M1, M2, F1, and F2. Hence, factor structures based on groups M1 and M2, and those based on groups F1 and F2, differ from each other only by random chance, whereas factor structures based on groups M1 and F1, and on groups M2 and F2, may differ from each other due to the sex of the respondent as well as random chance. Furthermore, a posteriori alterations of the a priori factor structure based on tests of M1 and F1 can be cross-validated with tests of M2 and F2, and a posteriori alterations to M2 and F2 can be cross-validated with M1 and F1.

For purposes of this analysis, as in other SDQ research, each of the eight SDQ factors is represented by four variables and each variable is the total score of two SDQ items that are designed to measure the same factor (see Marsh, 1986a; Marsh & Hocevar, 1985; Marsh, Barnes, et al. 1984). Thus, each of the analyses described later is based on one of four 32 x 32 covariance matrices derived from groups M1, M2, F1 or F2.

The Definition and Tests of the Confirmatory Factor Analyses (CFA) Models.

The definition of a priori models. The application of CFA for single groups, and the advantages of CFA over EFA for such purposes, are well known and will not be the focus of the present investigation (see Bagozzi, 1980; Joreskog, 1981; Joreskog & Sorbom, 1983; Long, 1983; Marsh, 1985b; Marsh & Hocevar, 1983; 1984; 1985; Pedhauzur, 1982; Wolfle, 1981). The parameter estimates from CFA are conceptually similar to those from conventional EFAs except that the researcher is able to fix or constrain elements in accordance to an a priori model to be tested. In the present investigation, three design matrices were used to define the a priori factor model: a 32 (variables) x 8 (factors) matrix of factor loadings; an 8 x 8 factor variance-covariance matrix which represents the relationships among the factors; and a 32 x 32 matrix of error/uniquenesses in which the diagonal elements are similar to one minus communality estimates in EFA and the off-diagonal elements, if estimated, represent correlated errors. The a priori model (see Table 2 in Results section) had a simple structure in that each measured variable was allowed to define only the factor it was designed to measure, and its loadings on all other variables were specified to be zero. One measured variable for each of the 8 SDQ factors was selected to be a reference indicator, and its loading was fixed to be 1.0. Values for the other 24 factor loadings, the 8 factor variances, the 28 factor covariances, and the 32 error/uniquenesses are estimated as part of the analysis. All other elements in the design matrices are specified to be zero.

Goodness of Fit in CFA. In CFA there are no well established guidelines for what minimal conditions constitute an adequate fit. The general approach is to:

1) examine estimated parameters in relation to the substantive, a priori structure (and also estimates outside the range of permissible values such as negative variance estimates);

2) evaluate the overall X^2 in terms of statistical significance, and compare this with values obtained from alternative models;

3) evaluate subjective indices of goodness of fit that give an indication of the proportion of variance that is explained by the model, and compare these indices from alternative models;

4) perhaps, if the a priori structure is not judged to adequately fit the data, to formulate alternative, a posteriori structures that fit the data better. However, when substantial changes are made to the original, a priori structure, the results should be interpreted cautiously and replicated with new data. For this reason, sets of responses by males and by

females were each divided into random halves in the present investigation in order to provide tests of the replicability and generalizability of a posteriori alterations in the original model.

All analyses in the present investigation were performed with the commercially available LISREL V program (Joreskog & Sorbom, 1981). LISREL V, after testing for identification, attempts to minimize a maximum likelihood function that is based on differences between the original (observed) and reproduced (predicted) covariance matrix, and provides an overall chi-square goodness-of-fit test. In contrast to traditional significance testing, the researcher often prefers a nonsignificant chi-square that indicates that the hypothesized model fits the observed data. There are, however, problems with this test. First, it is highly sensitive to departures from multivariate normality. Second, for large, complex problems (i.e., where there are many variables and many parameters to be estimated) the observed chi-square will nearly always be statistically significant even when there is a reasonably good fit to the data. Third, the chi-square test is strongly influenced by sample size so that a poor fit based upon a small sample size may result in a nonsignificant chi-square while a good fit based upon a large sample size may result in a significant X^2. Hence, most practical applications of CFA require a subjective evaluation of whether or not a statistically significant chi-square is small enough to constitute an adequate fit, and this is an important, unresolved issue in the use of CFA (see Bentler & Bonett, 1980; Fornell, 1983; Hoelter, 1983; Marsh & Balla, 1986; Marsh & Hocevar, 1984; 1985; Sobel & Bohrnstedt, 1985; for further discussion).

A number of alternative indices or indications of goodness-of-fit have been developed, and those to be used in the present investigation are defined in Table 1. None of the alternative measures of goodness-of-fit has been universally accepted, each has problems, and some of these problems are particularly relevant to this investigation. Ultimately, each of the alternative indices depends upon a subjective impression about what value reflects an adequate fit, and thus undermines some of the rigor of CFA.

Insert Table 1 About Here

Results For The A Priori Structure

The Fit of the A Priori Structure With No Invariance Constraints.

When the a priori structure is fit separately for each of the four groups (M1, M2, F1, F2), each estimated factor loading and factor variance (see Model 1 in Table 2 for results of M1 & F1) is large and statistically significant. These results provide support for the a priori model. While the X^2 for Model 1 is statistically significant for all four groups, other goodness-of-fit indices suggest that the fit may be reasonable (Table 3). However, even if a model with no invariance constraints (i.e., Model 1) is able to fit the data from each group separately, the results should not be interpreted to mean that the factor structure is invariant across the groups. Instead, the results indicate that the same pattern of parameter estimates are able to fit the data from each group, but not that these parameter estimates take on the same, or even similar, values for the different groups. Tests of factorial invariance require additional models that posit some or all of the parameters to be the same across different groups.

Insert Tables 2 & 3 About Here

Tests of the Factorial Invariance of the A Priori Structure.

Goodness-of-fit for models with invariance constraints. In tests of factorial invariance, some or all of the parameter estimates from Model 1 are constrained so that the estimates are the same across the groups being tested. In this sense, the goodness of fit of Model 1, without any invariance constraints, represents an absolute upper bound -- an optimum or a target -- for alternative models that impose equality constraints. For example, the X^2 for any tests of invariance across groups M1 and F1 must be at least 1849.2 (851.5 + 997.7), the sum of the X^2s for M1 and F1 for Model 1. No alternative models that require any or all parameters to be the same across the two groups could have a X^2 smaller than 1849.2, and the X^2 for such an alternative model would only approach 1849.2 to the extent that the parameter estimates in the unconstrained models for M1 and F1 are the same. Thus, to the extent that an alternative model with equality constraints is able to fit the data nearly as well as Model 1, then there is support for the alternative model and the invariance of the constrained parameters. Hence, the comparison of goodness of fits for Model 1 with those of subsequent models of factorial invariance is very important.

The statistical significance of the difference in X^2s for two nested models (see Bentler & Bonett, 1980: Long, 1983; Sobel & Bohrnstedt, 1985; for a discussion of nested models) can be evaluated relative to the difference in the df for the two models. For example, Model 2 differs from Model 1 in that the 24 factor loadings are required to be the same in the two groups being tested, and so the difference in df is 24. For the M1/F1 comparison the difference in X^2 for Models 1 and 2 is 27.5 (Table 3) and not statistically significant for df=24. Thus, these results provide strong support for the invariance of factor loadings in these two groups. Since this X^2 test is very powerful, it is also important to evaluate subjective indicators of goodness of fit even when the X^2_{diff} is statistically significant. The comparison of traditional goodness-of-fit indices for two nested models is useful, but Marsh (Marsh, 1985b; Marsh & Hocevar, 1985) also developed the target coefficient (TC; see Table 1) specifically for this purpose. For purposes of the present investigation, the TC scales the X^2 for the model with invariance constraints along a zero-to-one scale in which the zero-point is defined by the null model and the top of the scale is defined by the fit of the corresponding a priori model without any equality constraints (Model 1). Thus TC provides an estimate of the proportion of variance explained by the unconstrained Model 1 that can be explained by the constrained (nested) model.

The Invariance models to be tested. For purposes of the present investigation, different sets of parameters are specified to be invariant in Models 2-7. The minimum condition of factorial invariance is for the factor loadings to be invariant (Model 2), and Models 2-7 all require the 24 estimated factor loadings to be invariant. In Models 3-7 factor variances, factor correlations[1], error/uniquenesses and various combinations of these parameters are also constrained to be invariant across groups. For each model, four tests of factorial invariance are performed. Tests of factorial invariance across opposite-sex groups, M1/F1 and M2/F2, provide two separate tests of factorial invariance across responses by males and females. Tests of factorial invariance across same-sex groups, M1/M2 and F1/F2, provide two separate tests of factorial invariance across random split-halves that differ only by random chance.

Model 1 (No Invar) -- no invariance constraints are imposed and this model provides one basis of comparison for evaluating Models 2-6 as well as the ability of the unconstrained model to fit the data.

Model 2 (FL Invar) -- the 24 factor loadings (FL) are specified to be invariant across groups, and this model is taken to be the minimum condition of factorial invariance.

Model 3 (FL, Fcovar & Fvar Invar) -- the 24 factor loadings, the 24 factor covariances (Fcovar), and the 8 factor variances (Fvar) are specified to be invariant.

Model 4 (FL & Fcorr Invar) -- the 24 factor loadings and the 24 factor correlations (Fcorr; factor covariances that have been standardized -- see footnote 1) are specified to be invariant.

Model 5 (FL & UE Invar) -- the 24 factor loadings and the 32 error/uniquenesses (UEs) are specified to be invariant.

Model 6 (FL, Fcorr & UE Invar) -- the 24 factor loadings, the 24 factor correlations, and the 32 error/uniquenesses (UEs) are specified to be invariant.

Model 7 (Total Invar) -- all 92 parameter estimates are specified to be invariant.

It is important to realise that this set of models in no way exhausts all possibilities, and many other models could be hypothesized that are consistent with the theoretical nature of the study, or are suggested by the results of preliminary analyses.

Empirical Results For The Invariance of the A Priori Structure.

For the two same-sex sets of comparisons none of the invariance constraints imposed in Models 2 - 7 has any substantial effect on the X^2s (Table 3). For the two male samples, not even the test of total invariance (Model 7) differs significantly from Model 1 (i.e., the difference in X^2 of 126 is not statistically significant for df= 92). For the two female samples, the invariance of the structures is not met in a strict statistical sense, but the goodness-of-fit indicators show that the model of total invariance provides a reasonable fit; the subjective goodness of fit indicators (X^2/df, BBI, & TLI) are nearly the same for Models 1 and 7, and the TC for Model 7 is .984. These findings are reasonable since these two same-sex comparisons each involve comparisons across two random split halves that differ only by random chance. Nevertheless, the results do provide a strong demonstration of the replicability of the a priori factor structure designed to explain responses to the SDQ.

For the two opposite-sex comparisons, the various sets of invariance constraints have a somewhat larger effect on the X^2s. For Model 7 the X^2_{diff}s (210 and 206 with df=92) are statistically significant, and larger than for the same-sex comparisons. Hence, the hypothesis of total invariance may not be tenable. However, the invariance of factor loadings (Model 2) and factor loadings in combination with factor correlations (Model 4) are not statistically significant. The comparison of Models 3 and 4 suggests that the factor variances are not strictly invariant, whereas the comparison of Models 2 and 6 suggests that the uniqueness/errors are not strictly invariant. Thus Model 4, positing the invariance of factor loadings and factor correlations, appears to be the most restrictive model that can be unequivocally supported for the opposite-sex comparisons. Nevertheless, it should be noted that even Model 7 that posits total invariance provides a reasonable fit for these opposite-sex comparisons. This observation is based on the finding that the BBI, TLI, and the X^2/df ratios are nearly the same for Models 1 and 7, and that the TCs for Model 7 are .976 and .978. Hence, most of the variance that can be explained by the a priori structure with no invariance constraints can also be explained by the model of total factorial invariance for the opposite-sex comparisons. These findings support the generalizability of the a priori factor structure across responses by males and females.

Results For A Posteriori Structures.

Development of A Posteriori Structures.

The a priori structure provides a reasonable fit to the data, but is not acceptable on strictly statistical criteria. Since this situation is almost always the case for CFA studies, the inability to establish generalizable criteria of what constitutes an adequate goodness of fit is a serious problem. While the a priori structure provides a reasonable fit, a less restrictive structure, one in which nontarget loadings were estimated or error/uniquenesses were allowed to be correlated, might provide a better fit to the data. The determination of which parameter estimates should be freed has not been well established in the CFA literature. The best approach is to use the substantive nature of the data for developing alternative structures (e.g., the use of correlated errors is substantively reasonable for longitudinal panel data). However, in many applications, the decision is based on empirical results from a previous analyses of the same data. This sequential development of new structures based on tests of old structures that are tested with the same data has serious implications both for tests of statistical significance and for the replicability of the results to new data, and such problems are not unique to CFA (e.g., step-wise procedures in multiple regression have similar problems). When the a posteriori structure differs substantially from the a priori structure most researchers argue that such a model should be cross-validated with new data, but cross-validations are infrequent. Hence, one purpose of the present investigation is to demonstrate an application of such a cross-validation using random split-halves of the responses by males and by females.

Joreskog and Sorbom (1981) describe a modification index that is provided by LISREL V for each parameter that is fixed or constrained. The index is the lower-bound estimate of the expected decrease in the observed chi-square that would result if that particular parameter were freed, and they suggest that a modification index should be at least 5 before a model is modified. (A X^2 of 5 for df=1 is statistically significant at approximately p =.025.) As with the X^2, the modification index is substantially influenced by sample size, but it does provide a valuable tool for suggesting alternative structures. The modification indices for different parameters are not independent so that freeing two parameters with modification indices of five is unlikely to result in a change in chi-square of 10. Hence, Joreskog and Sorbom suggest that only one parameter should be modified and tested at a time. However, this suggestion is likely to result in unacceptable costs for practical applications in which there are many parameters with modification indices greater than 5.

For purposes of this study new parameters to be estimated in two a posteriori structures were selected on the basis of Model 1 which has no invariance constraints, so that the selection was independent of invariance constraints tested in subsequent models. Inspection of the modification indices for all four groups suggested that the fit would not be substantially improved by freeing nontarget loadings, but that freeing correlated errors might result in a substantial improvement[2]. The correlated errors to be estimated in the first a posteriori structure were selected on the basis of results from Model 1 as applied to groups M1 and F1. Correlated errors were estimated in the a posteriori structure when the corresponding modification index was greater than 5 for both M1 and F1, and a total of 15 (of 496) correlated errors were so selected. The second a posteriori structure was based on a similar procedure applied to groups M2 and F2, and 14 correlated errors were identified. In order to facilitate comparisons described below, one additional correlated error that had a modification index greater than 5 for one group but not the other was also selected so that a total of 15 correlated errors were estimated in each of the a posteriori structures. These two a posteriori structures, one based on M1 and F1 and one based on M2 and F2, were then fit to Models 1-7. Models 1-7 as applied here differ from those summarized earlier (also see Table 3) only in that: a) 15 correlated errors were estimated for each group; and b) Models 5-7 that constrain error/uniquenesses to be invariant also constrain the correlated errors to be invariant.

<u>The Goodness-of-fit For Models of the A Posteriori Structures.</u>

<u>Models with no invariance constraints.</u> Both a posteriori structures resulted in large and statistically significant improvements in X^2s for Model 1 (Tables 4 & 5) in comparison to the corresponding X^2s for the a priori structure (Table 3) for each of the four groups (differences in X^2s vary from 97 to 236 with df = 15 for the eight tests). As expected, the first a posteriori structure based on M1 and F1 produces a larger improvement in M1 and F1 than for M2 and F2 (Table 4), whereas the second a posteriori structure based on M2 and F2 produces a larger improvement for M2 and F2 than for M1 and F1 (Table 5). Nevertheless, these results demonstrate that correlated errors selected on the basis of one set of data

produce a substantial improvement in goodness of fit for another set of data. However, it is important to note that it is only the selection of parameters to be estimated that is cross-validated by these results and not the actual values of the parameters.

Insert Tables 4 & 5 About Here

Same-sex comparisons for models with invariance constraints. As described earlier the same-sex groups (M1 & M2, and F1 & F2) differ only by random chance, and results for the a priori structure indicated that all parameter estimates were reasonably invariant across these same-sex comparisons. However, tests of factorial invariance for the two a posteriori structures differ from those for the a priori structure. Since the correlated errors to be estimated in the first a posteriori structure were derived from M1 and F1, tests of invariance (Table 4) across M1 and M2, and across F1 and F2, constitute a rigorous test in which both the selection of parameters and the actual values of these parameters are cross-validated. Similarly, the correlated errors in the second a posteriori structure were derived from M2 and F2 so that tests of invariance (Table 5) across the same-sex groups constitute a second, equally rigorous cross-validation. As with the a priori structure, models that do not require the correlated errors to be invariant (Models 1 - 4 in Tables 5 & 6) are reasonably invariant. However, Models 5-7 that posit the invariance of the correlated errors do not support the invariance of these parameters in a strict statistical sense. In particular Models 5-7 have X^2s that are significantly larger than Model 1 for both a posteriori structures (Tables 5 & 6). Hence, these results suggest that the actual values estimated for parameters selected on the basis of one set of data do not cross-validate very well to results based on another set of data. Nevertheless, it should be noted that the goodness-of-fit indices for all such models of invariance across same-sex groups are reasonable, and differ only modestly from those based on Model 1 with no invariance constraints. Furthermore, even for the model of total invariance (Model 7) the TCs, which vary from .975 to .981, indicate that nearly all the variation explained by the unconstrained models can be explained in terms of the model of total factorial invariance.

Opposite-sex comparisons for models with invariance constraints. Results for opposite-sex comparisons for both a posteriori structures, as with the a priori structure, suggest that there is invariance of factor loadings (Model 2) and factor correlations (Model 4), but that factor variances and error/uniquenesses (including correlated errors) are not strictly invariant. Furthermore, the X^2_{diff}s for the a posteriori structures are substantially larger than those for the a priori structure for models that posit the invariance of correlated errors (Models 5 - 7). The interpretation of these results is complicated by the significant X^2_{diff}s for the a priori model and the lack of invariance of the correlated errors across random split-halves of the same group. Hence, the lack of fit of the most restrictive invariance models, apparently, is due in part to the lack of invariance across responses by males and females, and in part to the lack of replicability of a posteriori parameters selected on the basis of one set of data to another set of data. Nevertheless, this statistical lack of invariance also reflects the power of the X^2 test, as evidenced by the similarity of goodness-of-fit indicators for Models 1 and 7 and the large TCs, varying between .967 and .974, for Model 7. From a practical perspective, most of the variance that can be explained by the unconstrained models is explicable by even the most restrictive tests of factorial invariance.

Summary of Tests For the Two A Posteriori Structures.

A priori structures provided reasonably good fits to the data, but inspection of modification indices indicated that the addition of some correlated errors would improve the fit. Tests of two such a posteriori structures did provide better fits for all four groups, even when additional parameters selected on the basis of one set of data were tested with another set of data. However, the improvement in fit was larger for the random split-half groups used to select the additional parameters than for the random split-half groups used to cross-validate the selection. Furthermore, strict statistical tests that required the estimated values of the additional parameters to be the same across cross-validation groups were not satisfied. This suggests that some of the improvement due to the inclusion of additional parameters may have been illusory and may be explained by capitalizing on chance in the selection of additional parameters to be estimated.

Tests of factorial invariance across opposite-sex comparisons were complicated for the a posteriori comparisons. The X^2s for all opposite-sex comparisons, no matter which a posteriori model was used and what invariance constraints were imposed, were better than the corresponding X^2s for the a priori structures. As with the a priori structures, there was support of the invariance of factor loadings and factor correlations but support for the invariance of factor variances and error/uniquenesses was weaker. However, for tests that involved the invariance of the correlated errors, the X^2_{diff}s (between Model 1 and the tested model) were larger for the a posteriori structures than for the a priori structure. Thus, while the inclusion of correlated errors improves the goodness of fit for all models, the actual values of these correlated errors were not strictly invariant across the responses by males and females.

Summary, Discussion and Implications.

The Substantive Issue.

Historically self-concept researchers have generally been unable to identify salient factors in responses to self-concept instruments that were replicable across similar groups or that generalized across different groups. Results of the present investigation demonstrate that an a priori factor structure for responses to the SDQ was invariant across responses by random split-halves of the same group, and that the a priori structure was reasonably invariant across responses by males and by females. These results provide support for both the replicability and generalizability of the factor structure underlying responses to the SDQ.

Sex differences are frequently examined by self-concept researchers, but such comparisons are generally based on mean differences between groups that implicitly assume that the factor structures of responses by males and females are relatively invariant. Similarly, the comparison of mean responses across any groups (e.g., age groups, ethnic groups, experimental and control groups) implicitly assumes a reasonable invariance of the factor structure for responses by the groups, but the assumption is rarely examined. While a few researchers have used EFA to compare factors identified in responses by males and females, such comparisons are of limited utility for tests of factorial invariance. Hence, the demonstration of factorial invariance across responses by males and females is substantively important, provides further justification for the comparison of mean responses by males and females that has been examined in previous SDQ research, and provides a methodological demonstration that has wide applicability.

The Methodological Issue.

Analyses described in this section demonstrate how the invariance of a factor structure can be tested across different groups of subjects responding to the same set of stimuli. In this application, clear support was shown for the factorial invariance of parameters from an a priori structure across random split-halves of the same groups, but support was somewhat weaker for the opposite-sex comparisons. Based on the results from one set of data, additional parameters were used to define a posteriori structures that were cross-validated with a different set of data. While the a posteriori structures substantially improved the goodness of fit compared to the a priori structure, even when cross-validated with different data, the values of the estimated parameters were not strictly invariant across random split halves or across opposite-sex comparisons.

Important issues and misconceptions in the application of CFA to tests of factorial invariance were identified:

1) While CFA allows the researcher to rigorously define the model to be fit and to generate parameter estimates, indications of the ability of such models to fit the data are often subjective. Furthermore, conventional "rules of thumb" are not always appropriate, and their limitiation are not well understood. The comparison of goodness-of-fit indicators among alternative models, and models representing a null fit and logically constructed optimum fits, are more useful than attempts to interpret the absolute value of indicators for any one model according to rules of thumb (see Marsh & Balla, 1986; Sobel & Bohrnstedt, 1985 for further discussion).

2) When factor variances are estimated separately for each group, factor correlations (as opposed to factor covariances) can only be tested with a specially constructed model. Even though factor correlations are often the concern of researchers, support for the invariance of factor covariances does not imply that factor correlations are invariant, and rejection of the hypothesis of the invariance of the factor variance/covariance matrix does not imply that factor correlations are not invariant (see footnote 1).

3) Applications of CFA, including tests of factorial invariance, often posit an a priori structure which does not fit the data according to strict statistical criteria. An important unresolved question is whether additional a posteriori structures should be examined that better fit the data, even if the changes are based on empirical guidelines rather than the substantive issues. An emprical procedure for modifying the a priori structure in a way that substantially improved the fit was demonstrated. Parameters selected on the basis of one set of data improved the fit for another set of data. However, the actual values estimated for these parameters were not strictly invariant across either random split halves of the same groups, or across responses by males and females. Hence, some of the improved fit due to the inclusion of additional parameters was apparently illusory in that it could not be cross-validated. Furthermore, the inability to adequately summarize goodness of fit in CFA meant that the extent of this problem was difficult to gauge even when cross-validation samples were tested. Thus an important problem is the determination of the extent of bias -- improvements in fit that cannot be cross-validated -- due to using the tests of a priori structures to formulate a posteriori structures that are tested with the same data.

FOOTNOTES

1 -- Factor correlations depend on both factor covariances and factor variances, so that when factor covariances and factor variances are constrained to be equal, factor correlations are also equal. However, constraining factor covariances to be invariant but allowing factor variances to be free does not provide a test of the invariance of factor correlations. Marsh and Hocevar (1985) argued that it is often the invariance of factor correlations, rather than factor covariances, that is of theoretical interest. In order to test the invariance of factor correlations when factor variances are not invariant, factor variances are estimated in one diagonal matrix, while covariances among the factors are estimated in a second matrix where the diagonal values were set to 1's. In the specification of this LISREL model (e.g., Models 4 & 6 in Tables 3 - 5) there were: 32 y-variables (NY = 32); eight factors on the y-side (NE = 8); eight factors on the x-side (NK = 8); an 8 x 8 diagonal GAMMA matrix where factor variances were estimated; an 8 x 8 PHI matrix with the diagonal fixed to be 1's and estimated factor correlations in the off-diagonals; and a 32 x 32 THETA EPSILON matrix of error/uniquenesses (see Marsh & Hocevar, 1985 for further discussion).

2 -- Correlated errors are interpreted to mean that the error/uniqueness for one measured variable is more highly correlated with the error/uniqueness of another variable than can be explained in terms of the variables's mutual reliance on a common factor (if both variables are used to infer the same factor) or on the covariation between the factors inferred by each variable. This may occur, for example, if two items are nearly synonymous, if there is an unusual phrase that is common to two items, or if two variables share some other source of unique variation. The requirement that modification indices for the correlated errors be at least 5 for both M1 and F1 (or for M2 and F2) before the additional parameter was included in the first (or second) a posteriori structure provided a more conservative criterion than suggested by Joreskog and Sorbom (1981) and guaranteed that the selection of additional parameters had some generality across opposite-sex comparisons. Nevertheless, only 4 of the 15 correlated errors selected for the first a posteriori structure were also selected for the second a posteriori structure. Joreskog and Sorbom also suggested that only one parameter should be added at a time, since the modification indices for different parameters are not independent and the addition of one new parameter may affect the modification indices of other parameters. In the present investigation, all 15 correlated errors were added at one time (because of substantial

additional cost of adding one at a time), and 12 of 60 correlated errors (i.e., 15 for each of the four groups) were not statistically significant for the group used to select the correlated error. However, no additional correlated errors had a modification index greater than 5 for both M1 and F1 (or M2 and F2) in the first (second) a posteriori structure after the inclusion of the originally selected 15 correlated errors. It should also be noted that the modification index like the X^2, is substantially influenced by sample size so that the number of parameters with modifications indices greater than 5 will increase with the sample size. If the sample sizes are particularly large, it may be preferable to select a larger modification index as the criterion for inclusion of additional parameters so that the contribution is practically as well as statistically significant.

REFERENCES

Alwin, D. F., & Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. J. Jackson & E. F. Borgotta (Eds.), Factor analysis and measurement in sociological research: A multidimensional perspective. Beverly Hills, Calif.: Sage.

Bagozzi, R. P. (1980). Causal models in marketing. New York: Wiley.

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

Boersma, F.J. & Chapman, J.W. (1979). Student's Perception of Ability Scale Manual. University of Alberta. Edmonton, Canada.

Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. Review of Educational Research, 54, 427-456.

Dusek, J. B. & Flaherty, J. F. (1981). The development of self-concept during adolescent years. Monographs of the Society for Research in Child Development, 46, (4,Serial No. 191).

Everett, J. (1983). Factor Comparability as a means of determining the number of factors and their rotation, Multivariate Behavioral Research, 18, 197-218

Everett, J. & Entrekin, L. V. (1980). Factor comparability and the advantages o multiple group factor analysis. Multivariate Behavioral Research, 2, 165-180.

Fleming, J. S. & Courtney, B. E. (1984). The dimensionality of self-esteem: II: Hierarchical facet model for revised measurement scales. Journal of Personality and Social Psychology, 46, 404-421.

Fornell, C. (1983). Issues in the application of covariance structure analysis. Journal of Consumer Research, 9,443-48.

Gorsuch, R. I. (1974). Factor Analysis. Philadelphia: Saunders.

Harman, H. H. (1976). Modern Factor Analysis. Chicago: University of Chicago Press.

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

Hoelter, J. W. (1983) The analysis of covariance structures: Goodness- of-fit indices. Sociological Methods & Research, 11, 325-344.

Joreskog, K. G. (1981). Analysis of covariance structures. Scandinavian Journal of Statistics, 8, 65-92.

Joreskog, K. G. & Sorbom, D. (1981). LISREL V: Analysis of Linear Structural Relations By the Method of Maximum Likelihood. Chicago: International Educational Services.

Long, K. S. (1983). Confirmatory Factor Analysis. Beverly Hills, CA: Sage.

Marsh, H. W. (1985a). Age and sex effects in multiple dimensions of preadolescent self-concept. Australian Journal of Psychology, 197-204.

Marsh, H. W. (1985b). The structure of masculinity/femininity: An application of confirmatory factor analysis to higher-order factor structures and factorial invariance. Multivariate Behavioral Research, 20, 427-449.

Marsh, H. W. (1986a). The Self Description Questionnaire (SDQ): A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept: A test manual and a research monograph. Faculty of Education, University of Sydney, NSW Australia.

Marsh, H. W. (1986b). Verbal and math self-concepts: An internal/external frame of reference model. American Educational Research Journal, xxx-xxx.

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). The Self Description Questionnaire (SDQ): Age effects in the structure and level of self-concept for preadolescent children. Journal of Educational Psychology, 76, 940-956.

Marsh, H. W., Barnes, J., & Hocevar, D. (1985). Self-other agreement on multidimensional self-concept ratings: Factor analysis & multitrait-multimethod analysis. Journal of Personality and Social Psychology, 49, 1360-1377.

Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. Journal of Educational Measurement, 20, 231-248.

Marsh, H. W. & Hocevar, D. (1984). The factorial invariance of students' evaluations of college teaching. American Educational Research Journal, 21, 341-366.

Marsh, H. W. & Hocevar, D. (1985). The application of confirmatory factor analysis to the study of self-concept: First and higher order factor structures and their invariance across age groups. Psychological Bulletin, 97, 562-582.

Marsh, H. W., & Parker, J. W. (1984). Determinants of self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well. Journal of Personality and Social Psychology, 47, 213-231.

Marsh, H. W., Parker, J. W., & Smith, I. D. (1983). Preadolescent self-concept: Its relation to self-concept as inferred by teachers and to academic ability. British Journal of Educational Psychology, 53, 60-78.

Marsh, H. W., Relich, J. D. & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. Journal of Personality and Social Psychology, 45, 173-187.

Marsh, H. W. & Richards, G. (1986) The Outward Bound Bridging Course for low achieving high-school males: Effect on academic achievement and multidimensional self-concepts. (In Review). Faculty of Education, University of Sydney, Australia.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. Educational Psychologist, xxx-xxx.

Marsh, H. W., & Smith, I. D. (1982). Multitrait-multimethod analyses of two self-concept instruments. Journal of Educational Psychology, 74, 430-440.

Marsh, H. W., Smith, I. D. & Barnes, J. (1983). Multitrait-multimethod analyses of the Self Description Questionnaire: Student-teacher agreement on multidimensional ratings of student self-concept. American Educational Research Journal, 20, 333-357.

Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: Relationships with inferred self-concepts and academic achievement. Australian Journal of Psychology, 36, 367-386.

Marsh, H. W., Smith, I. D. & Barnes, J. (1985). Multidimensional self-concepts: Relationships with sex and academic achievement. Journal of Educational Psychology, 77, 581-596.

Marsh, H. W., Smith, I. D., Barnes, J. & Butler, S. (1983). Self-concept: Reliability, dimensionality, validity, and the measurement of change. Journal of Educational Psychology, 75, 772-790.

Pedhauzur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Rosenberg, M. (1965). Society and the adolescent child. Princeton: Princeton University Press.

Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Shavelson, R. J., & Marsh, H. W. (In Press). On the structure of self-concept. In R. Schwarzer (Ed.), Anxiety and cognitions. NJ Lawrence Earlbaum.

Soares, L. M. and Soares, A. T. (1977). The self-concept: Mini, maxi, multi? Paper presented at the 1977 Annual Meeting of the American Educational Research Association, New York.

Sobel, M. F. & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In Sociological Methodology 1985 (pp. 152-178). San Francisco: Jossey-Bass.

Wolfle, L. M. (1981). Causal models with unmeasured variables: An introduction to LISREL. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

Wylie, R. C. (1979). The self-concept. (Vol. 2) Lincoln: University of Nebraska Press.

TABLE 1

Description of Goodness-of-fit Indicators Used in The Present Investigation

Indicator	Description
X^2/df	The ratio of the chi-square (X^2) to the degrees-of-freedom (df).
BBI	The Bentler-Bonett Index is: 1 - N/T where N and T are the X^2s for the null model and model to be tested (see Bentler & Bonett, 1980).
TLI	The Tucker-Lewis index is: $[(N/df_n - T/df_t) / [(N/df_n - 1]$. where N/df_n and T/df_t are the X^2/df ratios for the null model and the model being tested (see Bentler & Bonett, 1980).
X^2_d	The difference between X^2 for models with invariance constraints and Model 1 with no invariance constraints.
TC	The Target Coefficient, a measure of the ability of a model with invariance constraints explain the covariation compared to the corresponding model with no invariance constraints, is defined as: (N - I)/(N - U) where N, I and U are the X^2s for the null model (N), the model with invariance constraints (I) and the corresponding model without any invariance constraints (U; Model 1 in the present application). TC varies between 0 and 1.

TABLE 2

Model 1 Parameter Estimates For Males (M) and Females (F)

		Factor Loading Matrix								Error/
		Phys	Appr	Peer	Prnt	Read	Math	Schl	Genl	Uniqueness
Phys1	M	1.00	0	0	0	0	0	0	0	1.26*
	F	1.00	0	0	0	0	0	0	0	1.98*
Phys2	M	1.19*	0	0	0	0	0	0	0	1.63*
	F	1.02*	0	0	0	0	0	0	0	2.16*
Phys3	M	1.07*	0	0	0	0	0	0	0	2.03*
	F	1.07*	0	0	0	0	0	0	0	1.76*
Phys4	M	.93*	0	0	0	0	0	0	0	2.10*
	F	1.11*	0	0	0	0	0	0	0	1.61*
Appr1	M	0	1.00	0	0	0	0	0	0	1.54*
	F	0	1.00	0	0	0	0	0	0	2.17*
Appr2	M	0	.86*	0	0	0	0	0	0	1.18*
	F	0	.84*	0	0	0	0	0	0	2.38*
Appr3	M	0	1.34*	0	0	0	0	0	0	1.53*
	F	0	1.32*	0	0	0	0	0	0	1.57*
Appr4	M	0	1.10*	0	0	0	0	0	0	1.11*
	F	0	1.32*	0	0	0	0	0	0	1.35*
Peer1	M	0	0	1.00	0	0	0	0	0	1.69*
	F	0	0	1.00	0	0	0	0	0	1.81*
Peer2	M	0	0	1.18*	0	0	0	0	0	1.34*
	F	0	0	1.16*	0	0	0	0	0	1.07*
Peer3	M	0	0	1.35*	0	0	0	0	0	1.81*
	F	0	0	1.23*	0	0	0	0	0	1.58*
Peer4	M	0	0	1.23*	0	0	0	0	0	1.50*
	F	0	0	1.20*	0	0	0	0	0	1.26*
Prnt1	M	0	0	0	1.00	0	0	0	0	.98*
	F	0	0	0	1.00	0	0	0	0	.94*
Prnt2	M	0	0	0	1.45*	0	0	0	0	.93*
	F	0	0	0	1.24*	0	0	0	0	1.39*
Prnt3	M	0	0	0	1.84*	0	0	0	0	1.35*
	F	0	0	0	2.18*	0	0	0	0	1.38*
Prnt4	M	0	0	0	1.68*	0	0	0	0	.85*
	F	0	0	0	1.94*	0	0	0	0	.62*
Read1	M	0	0	0	0	1.00	0	0	0	1.08*
	F	0	0	0	0	1.00	0	0	0	1.25*
Read2	M	0	0	0	0	1.04*	0	0	0	.90*
	F	0	0	0	0	1.11*	0	0	0	.92*
Read3	M	0	0	0	0	1.10*	0	0	0	1.21*
	F	0	0	0	0	1.16*	0	0	0	.65*
Read4	M	0	0	0	0	1.12*	0	0	0	1.34*
	F	0	0	0	0	1.13*	0	0	0	1.24*
Math1	M	0	0	0	0	0	1.00	0	0	2.11*
	F	0	0	0	0	0	1.00	0	0	1.24*
Math2	M	0	0	0	0	0	.91*	0	0	1.16*
	F	0	0	0	0	0	.92*	0	0	1.05*
Math3	M	0	0	0	0	0	1.04*	0	0	.87*
	F	0	0	0	0	0	.99*	0	0	1.10*
Math4	M	0	0	0	0	0	.99*	0	0	.74*
	F	0	0	0	0	0	.99*	0	0	1.05*

TABLE 2 continued

		Factor Loading Matrix								Error/
		Phys	Appr	Peer	Prnt	Read	Math	Schl	Genl	Uniqueness
Schl1	M	0	0	0	0	0	0	1.00	0	1.89*
	F	0	0	0	0	0	0	1.00	0	1.40*
Schl2	M	0	0	0	0	0	0	.84*	0	1.98*
	F	0	0	0	0	0	0	.79*	0	1.87*
Schl3	M	0	0	0	0	0	0	1.31*	0	1.72*
	F	0	0	0	0	0	0	1.46*	0	1.80*
Schl4	M	0	0	0	0	0	0	1.15*	0	1.19*
	F	0	0	0	0	0	0	1.24*	0	1.14*
Genl1	M	0	0	0	0	0	0	0	1.00	1.83*
	F	0	0	0	0	0	0	0	1.00	1.86*
Genl2	M	0	0	0	0	0	0	0	1.16*	1.11*
	F	0	0	0	0	0	0	0	1.57*	1.23*
Genl3	M	0	0	0	0	0	0	0	1.38*	.68*
	F	0	0	0	0	0	0	0	1.28*	1.30*
Genl4	M	0	0	0	0	0	0	0	1.04*	1.41*
	F	0	0	0	0	0	0	0	1.28*	1.63*

Factors Variance/Covariance Matrix (factor correlations above the main diagonal)

Factors		PHYS	APPR	PEER	PRNT	READ	MATH	SCHL	GENL
PHYS	M	3.35*	.43	.65	.33	.20	.23	.25	.74
	F	2.99*	.40	.56	.23	.20	.22	.27	.75
APPR	M	.77*	.94*	.55	.22	.20	.33	.34	.51
	F	.96*	1.93*	.30	.01	.02	.04	.16	.30
PEER	M	1.47*	.66*	1.53*	.36	.28	.27	.31	.63
	F	1.23*	.53*	1.63*	.29	.23	.28	.32	.81
PRNT	M	.41*	.15*	.31*	.49*	.36	.15	.25	.44
	F	.27*	.01	.25*	.46*	.12	.09	.15	.28
READ	M	.64*	.33	.60*	.43*	3.05*	.26	.55	.32
	F	.49*	.05	.41*	.12	1.91*	.26	.52	.35
MATH	M	.94*	.70*	.72*	.23	.99*	4.82*	.75	.36
	F	.73*	.10	.69*	.12	.70*	3.69*	.71	.36
SCHL	M	.77*	.56*	.65*	.30*	1.64*	2.81*	2.91*	.42
	F	.54*	.26	.49*	.12	.83*	1.58*	1.35*	.43
GENL	M	1.23*	.45*	.71*	.28*	.50*	.72*	.65*	.83*
	F	1.16*	.37*	.92*	.17*	.43*	.62*	.45*	.80*

* $p < .01$

<u>Note.</u> Parameters with Values of 0 and 1 were fixed and not estimated as part of the analysis, and so no tests of statistical significance were performed for these values. The four measured variables designed to measure each factor are the sums of responses to pairs of positively worded items. Factor correlations, standardized factor covariances, were derived from the factor covariances and are presented to facilitate interpretations.

TABLE 3

Goodness of Fit Indices for the CFA Models of Factorial Invariance Across Pairs of Groups (no correlated errors)

Model Description	X^2	df	X^2/df	BBI	TLI	X^2_d	df_d	TC
0) Null Model[a]								
M1/F1	10566.2	992	10.65	.00	.00	----	----	----
M2/F2	11313.4	992	11.40	.00	.00	----	----	----
M1/M2	10883.7	992	10.97	.00	.00	----	----	----
F1/F2	10995.9	992	11.08	.00	.00	----	----	----
1) No Invariance[a]								
M1/F1	1849.2	872	2.12	.82	.88	0	0	1.0
M2/F2	1888.1	872	2.17	.83	.89	0	0	1.0
M1/M2	1621.2	872	1.86	.85	.91	0	0	1.0
F1/F2	2116.6	872	2.42	.81	.86	0	0	1.0
2) Factor Loadings invariant								
M1/F1	1876.7	896	2.09	.82	.89	27.5	24	.997
M2/F2	1907.9	896	2.13	.83	.89	19.8	24	.998
M1/M2	1647.5	896	1.83	.85	.92	26.3	24	.997
F1/F2	2137.0	896	2.39	.81	.86	20.4	24	.997
3) Factor Loadings, factor variances and factor covariances invariant								
M1/F1	1969.0	932	2.11	.81	.88	119.8	60	.986
M2/F2	1981.4	932	2.13	.82	.89	93.3	60	.990
M1/M2	1693.1	932	1.82	.84	.92	71.9	60	.992
F1/F2	2194.3	932	2.35	.80	.87	77.7	60	.991
4) Factor Loadings and factor correlations invariant								
M1/F1	1926.4	924	2.08	.82	.89	77.2	52	.991
M2/F2	1959.6	924	2.12	.83	.89	71.5	52	.992
M1/M2	1682.8	924	1.82	.85	.92	61.6	52	.993
F1/F2	2187.6	924	2.37	.80	.86	71.0	52	.992
5) Factor Loadings and uniquenesses invariant								
M1/F1	1963.9	928	2.12	.81	.88	114.7	56	.987
M2/F2	2010.9	928	2.17	.82	.89	122.8	56	[illegible]
M1/M2	1710.7	928	1.84	.84	.92	89.5	56	[illegible]
F1/F2	2203.2	928	2.37	.80	.86	86.6	56	.990
6) Factor Loadings, factor correlations, and uniquenesses invariant								
M1/F1	2012.2	956	2.10	.81	.89	163.0	84	.981
M2/F2	2068.7	956	2.16	.82	.89	180.6	84	.981
M1/M2	1747.8	956	1.83	.84	.92	126.6	84	.986
F1/F2	2257.9	956	2.36	.79	.87	141.3	84	.984
7) Total Invariance								
M1/F1	2059.2	964	2.13	.81	.88	210.0	92	.976
M2/F2	2093.8	964	2.17	.82	.89	205.7	92	.978
M1/M2	1757.2	964	1.82	.84	.92	126.0	92	.986
F1/F2	2262.4	964	2.35	.79	.87	145.8	92	.984

Note. See Table 1 for a description of the goodness-of-fit indicators. The Null model hypothesizes complete independence of all measured variables and is used in computing other indicators (see Table 1). For Models 2 - 6, the X^2_d and df_d are the differences between the X^2 and df for the model being tested and Model 1 for which no invariance constraints were imposed.

[a] The X^2s for the four null models are 5505.8, 5060.4, 5377.9, 5935.5 with 496 df for M1, F1, M2 and F2 respectively, while the corresponding X^2s for Model 1 are 851.5, 997.7, 769.2, and 1118.9 with 436 df.

TABLE 4

Goodness of Fit Indices for the CFA Models of Factorial Invariance Across Pairs of Groups (With Correlated Errors Based on M1 and F1)

Model Description	X^2	df	X^2/df	BBI	TLI	X^2d	dfd	TC
1) No Invariance								
M1/F1	1461.1	842	1.74	.86	.92	0	0	1.0
M2/F2	1691.7	842	2.01	.85	.90	0	0	1.0
M1/M2	1388.3	842	1.65	.87	.94	0	0	1.0
F1/F2	1764.5	842	2.10	.84	.89	0	0	1.0
2) Factor Loadings invariant								
M1/F1	1487.3	866	1.72	.86	.93	26.2	24	.997
M2/F2	1707.1	866	1.97	.85	.91	15.4	24	.997
M1/M2	1411.5	866	1.63	.87	.94	23.2	24	.997
F1/F2	1788.3	866	2.07	.84	.89	23.8	24	.997
3) Factor Loadings, factor variances and factor covariances invariant								
M1/F1	1585.7	902	1.75	.85	.92	124.6	60	.986
M2/F2	1777.9	902	1.97	.83	.92	86.2	60	.991
M1/M2	1461.5	902	1.62	.87	.94	73.2	60	.992
F1/F2	1849.1	902	2.05	.83	.90	84.5	60	.991
4) Factor Loadings and factor correlations invariant								
M1/F1	1545.1	894	1.73	.85	.93	84.0	52	.991
M2/F2	1750.7	894	1.96	.85	.91	59.0	52	.994
M1/M2	1451.4	894	1.62	.87	.94	63.1	52	.993
F1/F2	1840.0	894	2.06	.83	.90	75.5	52	.992
5) Factor Loadings and uniquenesses invariant								
M1/F1	1660.4	913	1.82	.84	.92	199.3	71	.978
M2/F2	1861.3	913	2.04	.84	.90	169.6	71	.982
M1/M2	1539.4	913	1.69	.86	.93	151.1	71	.984
F1/F2	1909.6	913	2.09	.83	.89	145.1	71	.984
6) Factor Loadings, factor correlations, and uniquenesses invariant								
M1/F1	1714.1	941	1.82	.84	.92	253.0	99	.972
M2/F2	1914.1	941	2.03	.84	.90	222.4	99	.977
M1/M2	1581.3	941	1.68	.85	.93	193.0	99	.980
F1/F2	1961.6	941	2.08	.82	.89	197.1	99	.979
7) Total Invariance								
M1/F1	1764.2	949	1.86	.83	.91	303.1	107	.967
M2/F2	1941.0	949	2.05	.83	.90	249.3	107	.974
M1/M2	1590.0	949	1.68	.85	.93	201.7	107	.979
F1/F2	1970.1	949	2.08	.82	.89	205.6	107	.978

Note. The null models are the same as in Table 3. See note in Table 3 and Table 1 for a description of the goodness-of-fit indicators.

[a] The X^2s for the four individual group tests of Model 1 are 705.8, 755.3, 682.5, 1009.2 with 421 df for M1, F1, M2 and F2 respectively.

TABLE 5

Goodness of Fit Indices for the CFA Models of Factorial Invariance Across Pairs of Groups (With Correlated Errors Based on M2 and F2)

Model Description	X^2	df	X^2/df	BBI	TLI	$X^2 d$	dfd	TC
1) No Invariance[a]								
M1/F1	1571.4	842	1.86	.85	.91	0	0	1.0
M2/F2	1525.0	842	1.81	.87	.92	0	0	1.0
M1/M2	1399.4	842	1.66	.87	.93	0	0	1.0
F1/F2	1697.0	842	2.02	.85	.90	0	0	1.0
2) Factor Loadings invariant								
M1/F1	1597.1	866	1.84	.85	.91	25.7	24	.997
M2/F2	1542.3	866	1.78	.86	.93	17.3	24	.998
M1/M2	1417.8	866	1.64	.87	.94	18.4	24	.998
F1/F2	1723.8	866	1.99	.84	.90	26.8	24	.997
3) Factor Loadings, factor variances and factor covariances invariant								
M1/F1	1690.8	902	1.87	.84	.91	119.4	60	.987
M2/F2	1615.3	902	1.79	.86	.92	90.3	60	.987
M1/M2	1467.2	902	1.63	.87	.94	67.8	60	.993
F1/F2	1784.2	902	1.98	.84	.90	87.2	60	.991
4) Factor Loadings and factor correlations invariant								
M1/F1	1650.0	894	1.84	.84	.91	78.6	52	.991
M2/F2	1589.4	894	1.78	.86	.93	64.4	52	.993
M1/M2	1456.2	894	1.63	.87	.94	56.8	52	.994
F1/F2	1777.1	894	1.99	.84	.90	80.1	52	.991
5) Factor Loadings and uniquenesses invariant								
M1/F1	1719.8	913	1.88	.84	.91	148.4	71	.984
M2/F2	1727.3	913	1.89	.85	.91	202.3	71	.979
M1/M2	1534.4	913	1.68	.86	.93	135.0	71	.986
F1/F2	1869.9	913	2.04	.83	.90	172.9	71	.981
6) Factor Loadings, factor correlations, and uniquenesses invariant								
M1/F1	1771.2	941	1.88	.83	.91	199.8	99	.978
M2/F2	1786.1	941	1.90	.84	.91	261.1	99	.973
M1/M2	1574.4	941	1.67	.86	.93	175.0	99	.982
F1/F2	1924.1	941	2.04	.83	.90	227.1	99	.976
7) Total Invariance								
M1/F1	1819.7	949	1.92	.83	.91	248.0	107	.972
M2/F2	1812.7	949	1.91	.84	.91	287.7	107	.971
M1/M2	1583.5	949	1.67	.85	.93	184.1	107	.981
F1/F2	1929.3	949	2.03	.82	.90	232.3	107	.975

Note. The null models are the same as in Table 3. See note in Table 3 and Table 1 for a description of the goodness-of-fit indicators.

[a] The X^2s for the four individual group tests of Model 1 are 754.5, 816.9, 644.9, 880.1 with 421 df for M1, F1, M2 and F2 respectively.