Testing for Multigroup Invariance Using AMOS Graphics: A Road Less Traveled
Barbara M. Byrne
Available online: 19 Nov 2009

To cite this article: Barbara M. Byrne (2004): Testing for Multigroup Invariance Using AMOS Graphics: A Road Less Traveled, Structural Equation Modeling: A Multidisciplinary Journal, 11:2, 272-300

To link to this article: http://dx.doi.org/10.1207/s15328007sem1102_8

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages.
whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
The purpose of this article is to illustrate the steps involved in testing for multigroup invariance using Amos Graphics. Based on analysis of covariance (ANCOV) structures, 2 applications are demonstrated, each of which represents a different set of circumstances. Application 1 focuses on the equivalence of a measuring instrument and tests for its invariance across 3 teacher panels, given baseline models that are identical across groups. Application 2 centers on the equivalence of a postulated theoretical structure across adolescent boys and girls in light of baseline models that are differentially specified across groups. Taken together, these illustrated examples should be of substantial assistance to researchers interested in testing for multigroup invariance using the Amos program.

In substantive research that focuses on multigroup comparisons, it is typically assumed that the instrument of measurement is operating in exactly the same way, and that the underlying construct being measured has the same theoretical structure for each group under study. As evidenced from reviews of the literature, however, these two critically important assumptions are rarely, if ever, tested statistically. One approach to addressing this issue of equivalence is to test for the invariance of both the items and the factorial structure across groups using the analysis of covariance (ANCOV) structures. As such, one would test for the equivalence of all items designed to measure the construct underlying each subscale (i.e., factor), as well as relations among these underlying constructs. Operating within a parallel framework, this methodology can also be used to test for the multigroup invariance of a theoretical structure where the dimensionality of a psychological construct is of primary interest. For example, based on dual theories and their related empirical research, a researcher may wish to test for the invariance of a postulated four-factor structure versus a postulated two-factor structure of social self-concept (SC) for adolescents representing diverse cultural groups.
To date, almost all studies reported in the literature that address multigroup invariance based on the ANCOV structures have used either the LISREL (Jöreskog & Sörbom, 1996) or EQS (Bentler & Wu, 2002) programs. As a consequence, it has been customary to use the input file-based strategy associated with these programs. (For a detailed review and illustrated applications of this process for both LISREL and EQS, readers are referred to Byrne, 1998, 1994, respectively.) More recently, however, the less conventional graphical approach to ANCOV structures used by the Amos program (Arbuckle, 1999) has gained much popularity with researchers new to the application of this methodology. Although Amos incorporates a standard text-based interface (termed Amos Basic), its graphical interface (termed Amos Graphics) is the one more commonly used.

Despite this increased use of Amos, however, a review of the literature suggests its negligible use in testing for multigroup invariance. For example, a search of three journals in which multiple group comparisons are often presented yielded no evidence of tests for invariance using Amos; these were as follows: *Structural Equation Modeling* (Vol. 1, 1994–Vol. 10, 2003), *Journal of Cross-Cultural Psychology* (Vol. 27, 1996–Vol. 34, 2003), *Psychological Methods* (Vol. 1–Vol. 7, 2002). Given its rather unconventional approach to these analyses, this void likely derives from a lack of familiarity with the Amos strategy. The purpose of this article, then, is to provide some assistance in alleviating this difficulty. More specifically, based on ANCOV structures, and using the graphical interface of the Amos 4.0 program, the process of testing for equivalence in light of two different scenarios is illustrated: (a) The hypothesized multigroup model is identically specified across groups, and (b) the hypothesized multigroup model is differentially specified across groups. Although, conceptually, the analytic approach is similar across these two perspectives, technically, it must necessarily differ when using Amos Graphics. Both situations are addressed in this article. In Application 1, testing for the invariance of a measuring instrument for which the specified factorial structure of the measure is identical across groups is demonstrated. In Application 2, testing for the invariance of a theoretical construct for which the pattern of factor loadings differs across groups is demonstrated.

**TESTING FOR MULTIGROUP INVARIANCE: THE GENERAL PROCEDURE**

In testing for equivalencies across groups, sets of parameters are put to the test in a logically ordered and increasingly restrictive fashion, depending on the model and hypotheses to be tested. In the case of tests for the invariance of a measuring instrument and/or the invariance of a theoretical construct, the two applications to be illustrated in this article, only the factor-loading regression paths and the factor covariances are of interest. Except in particular instances when, for example, it
might be of interest to test for the equivalent reliability of an assessment measure across groups (see, e.g., Byrne, 1988), the equality of error variances and covariances is probably of least importance (Bentler, 2004). Indeed, it is now widely accepted that testing for the invariance of these error parameters represents an overly restrictive test of the data.

In the Jöreskog tradition, tests of hypotheses related to group invariance typically begin with scrutiny of the measurement model. In particular, the pattern of factor loadings for each observed measure is tested for its equivalence across the groups. Once it is known which observed measures are group invariant, these parameters are constrained equal while subsequent tests of the structural parameters are conducted. As each new set of parameters is tested, those known to be group invariant are constrained equal across groups. Given the univariate approach to the testing of these hypotheses, as implemented in the Amos program—compared with, for example, the multivariate approach used in the EQS program (Bentler & Wu, 2002)—this orderly sequence of analytic steps is both necessary and strongly recommended.

As a prerequisite to testing for factorial invariance, it is customary to consider a baseline model, which is estimated for each group separately. This baseline model represents one that best fits the data from the perspectives of both parsimony and substantive meaningfulness. However, measuring instruments are often group specific in the way they operate, and thus, baseline models are not expected to be completely identical across groups. For example, whereas the baseline model for one group might include cross-loadings and/or error covariances, this may not be so for other groups under study. A priori knowledge of such group differences is critical to the application of invariance-testing procedures. Although the bulk of the literature suggests that the number of factors must be equivalent across groups before further tests of invariance can be conducted, this strategy represents a logical starting point only, and is not a necessary condition. Indeed, only the similarly specified parameters within the same factor need be equated (see, e.g., Byrne, Shavelson, & Muthén, 1989; Werts, Rock, Linn, & Jöreskog, 1976).

Because the estimation of baseline models involves no between-group constraints, the data can be analyzed separately for each group. However, in testing for invariance, equality constraints are imposed on particular parameters, and thus, the data for all groups must be analyzed simultaneously to obtain efficient estimates (Bentler, 2004; Jöreskog & Sörbom, 1996); the pattern of fixed and free parameters...
ters nonetheless remains consistent with the baseline model specification for each group. We turn now to the example applications of interest in this article.

APPLICATION 1

In this first application, hypotheses related to the invariance of a single measuring instrument are tested across three groups of teachers. Specifically, we test for equivalency of a 20-item adaptation of the Maslach Burnout Inventory (MBI; Maslach & Jackson, 1986) across elementary (n = 1,159), intermediate (n = 388), and secondary (n = 1,384) teachers.

The Hypothesized Model

The model used in this first application is taken from a study by Byrne (1993) in which the initial task was to test for the validity of the MBI, a 22-item instrument designed to measure three dimensions of burnout—emotional exhaustion, depersonalization, and personal accomplishment. Based on consistent findings that Item 12 (designed to measure Personal Accomplishment) and Item 16 (designed to measure Emotional Exhaustion) were problematic in fitting the model to the data for elementary, intermediate, and secondary school teachers, a modified version of the instrument that excluded these items was proposed. The hypothesized model to be tested here is based on this 20-item adaptation of the MBI.3

The Baseline Models

In testing for the validity of the 20-item MBI model for each teacher group, findings were consistent in revealing exceptionally large correlated errors between Items 1 and 2 and between Items 10 and 11. Scrutiny of the content for each of these items revealed evidence of substantial overlap between each of these item pairs, a situation that can trigger error covariances. In light of this substantive justification, these error terms were subsequently specified as free parameters in the model for each teacher group. A final model that reflected these modifications was fully cross-validated for independent samples of elementary, intermediate, and secondary teachers.

This testing for a baseline model, then, yielded one that was identically specified for each of the three teaching panels. However, it is important to emphasize that just because the revised model was similarly specified for each teacher group, it in no

---

3For a more detailed account of analyses leading up to the 20-item model, readers are referred to the original article (Byrne, 1993), or to Byrne (2001).
way guarantees the equivalence of item measurements and underlying theoretical structure across teacher groups; these hypotheses must be tested statistically.

The hypothesized model under test in this example represents the revised 20-item MBI structure, together with the addition of two error covariances, as schematically depicted in Figure 1. At issue is the extent to which this modified version of the MBI is equivalent across elementary, intermediate, and secondary school teachers.

FIGURE 1  Baseline model of revised 20-item MBI structure for elementary, intermediate, and secondary teachers. From Byrne, 2001. Reprinted with permission.
TESTING FOR MULTIGROUP INVARIANCE
USING AMOS GRAPHICS

When working with ANCOV structures that involve multiple groups, the data related to each group must, of course, be made known to the program. Typically, for most structural equation modeling (SEM) programs, the data reside in some external file, the location of which is specified in an input file. Although no input file is used with Amos Graphics, both the name of each group and the location of its data file must be conveyed to the program prior to the analyses. This procedure is accomplished via the Manage Groups dialog box, which is made available either by pulling down the Model-Fit menu and selecting the “Manage Groups” option, or by using the Manage Groups icon. To begin, we click on “New” in the Manage Groups dialog box, which is shown in Figure 2. Each click will yield the name “Group,” along with an assigned number. In the case of Figure 2, the group number (3) pertains to secondary teachers; this name change is invoked simply by typing over the former name (see Figure 3).

Once the group names have been established, the next task is to identify a data file for each, which is accomplished through activation of the Data File dialog box. This information is made available either by clicking on the Data File icon, or by pulling down the File menu and selecting the “Data Files” option. The Data File dialog box for this application is shown in Figure 4.

Specification of a multigroup model, using Amos Graphics, is guided by the default rule that, unless explicitly declared otherwise, all groups in the analysis have an identical path diagram structure. As such, a model structure needs only to be drawn for the first group; all other groups will have the same structure by default. Given that all three groups in this first application have the same baseline model, this default rule poses no problem.

Testing for invariance necessarily entails a multistep process. However, when the analyses involve more than two groups, and findings reveal evidence of noninvariance, the number of steps required in identifying the source of such noninvariance can increase substantially. In the interest of clarity, given that this example involves three groups, each step of the process has been identified accordingly.

Step 1: Testing for the Validity of the Hypothesized Model Across Elementary, Intermediate, and Secondary Teachers

As a preliminary step in testing for invariance across groups, we test for the validity of MBI structure as best represented by the hypothesized three-factor structure...


shown in Figure 1. Given that this test of model fit was previously conducted in the process of determining the baseline models, readers may wonder why it is necessary to repeat the process. There are two important reasons for doing so. First, whereas the former tests were conducted for each group separately, tests for the validity of factorial structure in this instance are conducted across the three groups simultaneously. In other words, parameters are estimated for all three groups at the same time. Second, in testing for invariance using the Amos program, as with other SEM programs, the fit of this simultaneously estimated model can provide the baseline value against which all subsequently specified models are compared. In contrast to single-group analyses, however, this multigroup analysis yields only one set of fit statistics for overall model fit. Given that chi-square statistics, together with their degrees of freedom, are summative, the overall chi-square value for the multigroup model should equal the sum of the chi-square values obtained when the baseline model is tested separately for each group of teachers. This multigroup model reflects the extent to which the MBI structure fits the data when no cross-group constraints are imposed.

Model assessment. Goodness-of-fit statistics related to this three-group unconstrained model (Model 1) are reported in Table 1. The chi-square value of 2243.21, with 495 df, provides the baseline value against which subsequent tests for invariance may be compared. Comparative fit index (CFI) and root mean squared error of approximation (RMSEA) values of .93 and .04, respectively, indicated that the hypothesized three-factor model of MBI structure, although somewhat less than the recommended cutoff criterion of .95 recommended by Hu and Bentler (1999), still represented a relatively good fit across the three panels of teachers. Accordingly, we now proceed in testing for the invariance of the revised 20-item MBI across groups. (Readers interested in detailed information provided in the Amos output files are referred to Byrne, 2001.)

Step 2: Testing for Invariance of the Fully Constrained Model Across Elementary, Intermediate, and Secondary Teachers

Prior to testing for the equality of sets of parameters, as outlined earlier, it is always worthwhile to test for the possibility that a fully constrained model is invariant across groups. Regarding this application, this would mean specification of a model in which all factor loadings, all factor variances, all factor covariances, and the two error covariances are constrained equal across elementary, intermediate, and secondary teachers. Although, in general, testing for the equality of error variances across groups is considered to be excessively stringent, testing related to the error covariances specified in this context is well justified both statistically and substantively. This decision was based on the fact that, for each teacher group, the two error covariances were found to be excessively large. Scrutiny of the items as-
<table>
<thead>
<tr>
<th>Model Description</th>
<th>Groups</th>
<th>Comparative Model</th>
<th>χ²</th>
<th>df</th>
<th>∆χ²</th>
<th>∆df</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hypothesized model (Model 1)</td>
<td>Elementary, intermediate, secondary</td>
<td>Model 1</td>
<td>2243.21</td>
<td>495</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Factor loadings, variances, and covariances, plus error covariances constrained equal</td>
<td>Elementary, intermediate, secondary</td>
<td>Model 1</td>
<td>2344.75</td>
<td>545</td>
<td>101.54</td>
<td>50</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>3. Hypothesized model (Model 1a)</td>
<td>Elementary, intermediate</td>
<td>—</td>
<td>1246.24</td>
<td>330</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4. Factor loadings, variances, and covariances, plus error covariances constrained equal</td>
<td>Elementary, intermediate</td>
<td>Model 1a</td>
<td>1282.18</td>
<td>355</td>
<td>35.94</td>
<td>25</td>
<td>ns</td>
</tr>
<tr>
<td>5. Hypothesized model</td>
<td>Elementary, secondary</td>
<td>Model 1b</td>
<td>1990.06</td>
<td>330</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6. Factor loadings constrained equal</td>
<td>Elementary, secondary</td>
<td>Model 1b</td>
<td>2023.20</td>
<td>347</td>
<td>33.14</td>
<td>17</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>7. Factor loadings on EE constrained equal (Model 2)</td>
<td>Elementary, secondary</td>
<td>Model 1b</td>
<td>1992.40</td>
<td>337</td>
<td>2.34</td>
<td>7</td>
<td>ns</td>
</tr>
<tr>
<td>8. Model 2 with factor loadings on DP constrained equal</td>
<td>Elementary, secondary</td>
<td>Model 2</td>
<td>2005.52</td>
<td>341</td>
<td>13.12</td>
<td>4</td>
<td>p &lt; .01</td>
</tr>
<tr>
<td>9. Model 2 with factor loadings of Item 10 on DP constrained equal</td>
<td>Elementary, secondary</td>
<td>Model 2</td>
<td>1992.60</td>
<td>338</td>
<td>0.20</td>
<td>1</td>
<td>ns</td>
</tr>
<tr>
<td>10. Model 2 with factor loadings of Items 10 and 11 on DP constrained equal</td>
<td>Elementary, secondary</td>
<td>Model 2</td>
<td>2000.56</td>
<td>339</td>
<td>8.16</td>
<td>2</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>11. Model 2 with factor loadings of Items 10 and 15 on DP constrained equal</td>
<td>Elementary, secondary</td>
<td>Model 2</td>
<td>1994.75</td>
<td>339</td>
<td>2.35</td>
<td>2</td>
<td>ns</td>
</tr>
<tr>
<td>12. Model 2 with factor loadings of Items 10, 15, and 22 on DP constrained equal (Model 3)</td>
<td>Elementary, secondary</td>
<td>Model 2</td>
<td>1995.02</td>
<td>340</td>
<td>2.62</td>
<td>3</td>
<td>ns</td>
</tr>
<tr>
<td>13. Model 3 with factor loadings on PA constrained equal</td>
<td>Elementary, secondary</td>
<td>Model 3</td>
<td>2012.72</td>
<td>346</td>
<td>17.70</td>
<td>6</td>
<td>p &lt; .01</td>
</tr>
</tbody>
</table>

Note. ∆χ² = difference in chi-square values between models; ∆df = difference in number of degrees of freedom between models; EE = Emotional Exhaustion; DP = Depersonalization; PA = Personal Accomplishment.
associated with these error terms revealed highly overlapping content across each aberrant pair of items. As noted earlier, such redundancy can reflect itself in the form of error covariation. Based on this substantive rationale, then, it seems prudent to ascertain whether the two error covariance parameters hold across teaching panels. This fully constrained model is shown in Figure 5.

In SEM, testing for the invariance of parameters across groups is accomplished by placing constraints on particular parameters or, in other words, specifying par-

![Figure 5](image-url)
ticular parameters to be invariant (i.e., equivalent) across groups. In Amos Graphics, constraints are specified through a labeling mechanism whereby each parameter to be held equal across groups is given a label. Thus, in analyzing multigroup models, any parameters that are unlabeled will be freely estimated, thereby taking on different values across groups. To initiate this labeling process, we first activate the hypothesized model shown in Figure 1, and then, using the Object Properties dialog box, begin the task of labeling the parameters of interest. Shown in Figure 6, is the menu yielded through a right-click of the mouse, from which Object Properties is selected. More specifically, the Object Properties dialog box was triggered by first clicking on the second factor-loading path (which then becomes highlighted in red), and then clicking on the right mouse button. Figure 7 illustrates the
labeling of the first estimated parameter (p2) which represents the factor loading of MBI Item 2 on the factor, Emotional Exhaustion.

Let’s turn back again to Figure 5, where three aspects of the path diagram are worthy of particular note. First, selected labeling of parameters is purely arbitrary. In this case, the factor-loading regression paths are labeled as p, the factor variances as v_, the factor covariances as c_, and the error covariances as v_err. Thus, for example, v_ee represents the variance of the Emotional Exhaustion factor, c_eepa represents the covariance between the factors of Emotional Exhaustion and Personal Accomplishment, v_err1011 represents the covariance between the error terms associated with Items 10 and 11, and so on. Second, you will note that the value of 1.00, assigned to the first of each congeneric set of indicator variables (for purposes of statistical identification), remains as such and has not be relabeled with a “p”; given that this parameter is already constrained to equal 1.00, its value will be constant across the three groups. Finally, the somewhat erratic labeling of the factor-loading paths is a function of the automated labeling process provided by the Amos program and would appear to be related to the restricted space allotment assigned to each parameter. Although, technically, it is possible to shift labels to a more appropriate location using the Move Parameters tool, this transition does not seem to work well when there are several labeled parameters located in close proximity to one another, as is the case here.

**Model assessment.** Goodness-of-fit statistics related to this constrained three-group model are presented as the second entry in Table 1. In testing for the invariance of this constrained model, we compare its chi-square value of 2344.75 (545 df) with that for the initial model (Model 1) in which no equality constraints were imposed, $\chi^2(495) = 2243.21$. As with single-group applications, when models are nested, this difference in chi-square values (in large samples) is distributed as $\chi^2$, with degrees of freedom equal to the difference in degrees of freedom. Given that the constrained model is nested within the initial model, we use this comparative procedure here. This comparison yields a chi-square difference ($\Delta\chi^2$) value of 101.54 with 50 df, which is statistically significant ($p < .05$). Provided with this information, we now know that some equality constraints do not hold across the three teacher groups.

**Step 3: Testing for Invariance of Fully Constrained Model Across Elementary and Intermediate Teachers**

Given that we are working with three groups in this application, one approach to this series of analyses is to determine, first, if the constrained model is possibly invariant across two of the three groups of teachers. To this end, the hypothesized model (Model 1; no equality constraints) is once again estimated to establish a comparative base. However, in contrast to the previously estimated model, specifi-
cation relates only to elementary and intermediate teachers. This multigroup model has been labeled as Model 1a in Table 1 to distinguish it from the original multigroup model that included the three teacher groups.

In Amos Graphics, any change in the number of groups to be included in an analysis is accomplished via the Manage Groups dialog box. Consistent with Model 1a, secondary school teachers were excluded from the analysis by a simple highlighting and deletion process.

Model assessment. Turning again to Table 1, we focus on the third and fourth table entries. Entry 3 reports goodness-of-fit statistics related to the testing of Model 1a in which no equality constraints were imposed across elementary and intermediate teachers. Entry 4 reports on its comparison with the related constrained model. As you will readily observe, this comparison of models was not statistically significant, thereby indicating that all factor loadings, variances, and covariances, in addition to the two error covariances, are invariant across elementary and intermediate teachers. From these findings we can conclude that any inequality of parameters across the three groups of teachers must logically lie between secondary teachers and their elementary–intermediate school counterparts.

In this study, the category of intermediate teachers represented teachers working within the framework of an intermediate (or middle) school; all taught students in Grades 7 and 8. However, the category of elementary teachers also included teachers of Grades 7 and 8. Within this context, then, we can consider intermediate teachers under the broader rubric of elementary (as opposed to secondary) school teachers. In light of this consideration, together with the findings of equivalent factorial structure of the MBI across elementary and intermediate teachers, it would seem reasonable to treat these two teacher groups as one. Hence, for the remaining analyses, data for elementary and intermediate teachers were merged and treated as a single group. In the interest of simplicity, and to minimize possible confusion in the use of a combined group name, the label for this group is revised from elementary–intermediate teachers to elementary teachers is revised.

Because Amos Graphics provides no internal mechanism for merging data, this process must be accomplished using alternative means. However, once structured and imported to Amos, the newly merged data set is easily incorporated for use in running the remaining analyses. For example, let us say we save the merged data as a text file and label it as elemint.txt (consistent with the other data files shown in Figure 4). In preparation for the analyses to follow, we access the Manage Groups dialog box (see Figures 2 and 3) and do the following: (a) erase both the elementary and intermediate teacher groups by highlighting and then deleting each; (b) click on the “New” button, which will yield “Group number 2”; (c) replace Group number 2 with “Elemint Teachers.” The final task is to link this newly merged group to a data set. As such, we access the Data Files dialog box (see Figure 4). With Elemint Teachers highlighted, click on the “File Name” button, which in
Step 4: Testing for Invariance of Factor Loadings Across Elementary and Secondary Teachers

In testing for multigroup invariance, it is best to establish a logically organized strategy. The general scheme to be followed here is that we test first for the invariance of all factor loadings (i.e., all elements of the factor-loading matrix). Given findings of noninvariance at this level, we then proceed to test for the invariance of all factor loadings in each subscale (i.e., all loadings related to the one particular factor), separately. Given evidence of noninvariance at the subscale level, we then test for the invariance of each factor loading (related to the factor in question) separately. Of import in this process is that, as factor-loading parameters are found to be invariant across groups, their specified equality constraints are maintained, cumulatively, throughout the remainder of the invariance-testing process.

Turning to our task of testing for the invariance of the MBI across elementary and secondary teachers, then, we begin first by establishing goodness of fit for a multigroup unconstrained baseline model, which we label as Model 1b to distinguish it from Model 1 involving elementary, intermediate, and secondary teachers, and Model 1a involving only elementary and intermediate teachers. Nonetheless, it is important to emphasize that, although the groups tested varied, the hypothesized three-factor model of MBI structure remained the same across Models 1, 1a, and 1b. As shown in Entry 5 of Table 1, the testing of Model 1b yielded a $\chi^2(330)$ value of 1990.06.

We turn next to testing for invariant factor loadings related to all three factors — Emotional Exhaustion, Depersonalization, and Personal Accomplishment. In using Amos Graphics, this change in model specification necessarily requires a modification of the fully constrained model shown in Figure 5. Specifically, all labels, except those representing equality-constrained parameters, must be erased. As such, all parameter labels, except those associated with factor loadings, are removed. This task is easily accomplished by clicking on each label (to be deleted), right-clicking to trigger the Object Properties dialog box, and then deleting the label listed in the parameter rectangle (of the dialog box).

**Model assessment.** Turning to Entry 6 in Table 1, we see that testing of this model yielded a $\chi^2$ value of 2023.20 with 347 df; comparison with Model 1b yielded a $\Delta \chi^2$ value of 33.14 with 17 df, which was statistically significant ($p < .05$). To pinpoint the nonequivalent factor loadings, the next step entailed testing for invariance relative to each factor separately. Accordingly, the labels for Factor
2 (Depersonalization) and Factor 3 (Personal Accomplishment) were erased, leaving only those associated with Factor 1 (Emotional Exhaustion). Results related to this test of invariance, as indicated in Table 1 (Entry 7), revealed all factor loadings to be equivalent across elementary and secondary teachers; that is to say, comparison of this model with Model 1b revealed no significant differences.5

Having established the equivalence of factor loadings related to Emotional Exhaustion, these constraints were held in place while proceeding next to test for the invariance of the factor loadings on the Depersonalization factor. As such, model specification in the next round of analyses, using Amos Graphics, would have all factor loadings labeled for both Factors 1 and 2. Turning once again to Table 1 (Entry 8), we see that this test for invariance was statistically significant ($p < .01$), thereby signaling some discrepancy in the measurement of Depersonalization between elementary teachers and their secondary school peers. In other words, one or more items in this subscale were found to be noninvariant across the groups; the task then was to pinpoint these noninvariant items.

This search began by removing all labels associated with Factor 2, except the one associated with the first estimated parameter (Item 10; $p_{10}$). It is perhaps important to reconfirm that, given the invariance of all items in the Emotional Exhaustion subscale, the labeling of all factor loadings linked to Factor 1 remained intact (hence the description as Model 2) with the factor loading of Item 10 on Depersonalization constrained to be equal across groups. For clarification of this labeling process, specification of this revised model is displayed graphically in Figure 8.

As reported in Table 1 (Entry 9), the test for invariance related to this first factor-loading parameter (Item 10) was nonsignificant, thereby indicating its equality across groups. This orderly process of testing for the invariance of parameters is continued until all targeted parameters have been tested. It is perhaps worthwhile to reemphasize that, as parameters are found to be invariant, equality constraints related to these parameters are cumulatively held in place, thereby providing a very rigorous test of equality across groups. As indicated in Table 1, results from this series of tests bearing on Factor 2 revealed only the factor loading associated with Item 11 to be noninvariant across groups. For purposes of further comparisons involving the Personal Accomplishment factor (as explained earlier for comparisons related to the Depersonalization factor), this model is labeled as Model 3.

Having now established invariance related to the factor loadings for Factors 1 and 2, the next logical step was to test for the invariance of all factor loadings related to Personal Accomplishment. As such, this model would graphically reveal labels as-

---

5For purposes of simplicity in describing subsequent invariance models to be tested, this model is labeled as Model 2. As such, it eliminates the need to specify these constraints repeatedly for each of the models that follow.
associated with (a) all factor loadings associated with Emotional Exhaustion (p2–p8); (b) only those associated with Items 10 (p10), 15 (p12), and 22 (p13) for Depersonalization; and (c) all factor loadings associated with Personal Accomplishment (p15–p20). As indicated by their explicit labeling, factor loadings associated with Emotional Exhaustion and Depersonalization have been found to be group invariant; those associated with Personal Accomplishment are yet to be tested.

Results related to testing for the invariance of factor loadings related to Personal Accomplishment, as can be seen in Table 1 (Entry 13), indicated that certain items
in this scale were not invariant across elementary and secondary teachers. The course of action to be followed at this point, then, was to pinpoint these non-invariant items. As previously demonstrated with the testing for invariance of Factor 2 (Depersonalization), this process would be continued until all targeted parameters of interest have been tested. Once all tests for the invariance of factor loadings have been completed, one would next test for the equivalence of the two error covariances. Finally, once all tests for invariance related to the measurement model have been completed, one would turn next to the structural model in testing for the equivalence of the three factor covariances across the three panels of teachers. Due to the necessary limitation of space in describing these procedures, the latter tests of invariance are not detailed here. However, hopefully, sufficient material has been presented in this application to give readers an adequate understanding of the ordered process involved.

APPLICATION 2

This second multigroup application tests hypotheses related to the equivalence of SC measurement and structure across gender for high school adolescents. In particular, it seeks to determine if multidimensional facets of SC (general, academic, English, mathematics), as measured by subscale scores from multiple-assessment instruments, are equivalent across male \((n = 412)\) and female \((n = 420)\) adolescents.

The Hypothesized Model

The model on which this application is based (Byrne & Shavelson, 1987) derives from an earlier construct validity study of adolescent SC (Byrne & Shavelson, 1986) that argued for a four-factor structure composed of general SC (GSC), academic SC (ASC), English SC (ESC), and mathematics SC (MSC), with the four constructs assumed to be intercorrelated.

Except for ASC, each SC dimension is measured by three independent measures, each of which represents the related subscale score from one of four instruments: the Self Description Questionnaire–III (SDQ–III; Marsh, 1992), the Affective Perception Inventory (API; Soares & Soares, 1979), the Self-Concept of Ability Scale (SCA; Brookover, 1962), and the Self-Esteem Scale (SES; Rosenberg, 1989). At issue in the Byrne and Shavelson (1987) study was whether the structure of SC and the instruments used in measuring components of this structure were equivalent across adolescent boys and girls.

---

6Because preliminary analyses revealed the ASC subscale of the API to be a weak measure of this construct, it was deleted from all subsequent analyses.
The Baseline Models

In fitting the baseline model for each gender, Byrne and Shavelson (1987) reported a substantial drop in chi-square when the SDQESC of the SDQ–III was free to cross-load onto the GSC factor. Moreover, for boys only, the MSC subscale of the SDQ–III (SDQMSC) was allowed to cross-load onto the ESC factor. Finally, based on their substantial size and substantive reasonableness, three error covariances were included in the baseline model for both boys and girls. These baseline models were considered optimal in representing the data for adolescent boys, $\chi^2(33) = 86.07$ (CFI = .98; RMSEA = .063) and girls, $\chi^2(34) = 77.30$ (CFI = .99; RMSEA = .055), and are shown schematically in Figure 9.

Of primary interest regarding tests for invariance in this example, then, is the extent to which a four-factor structure of SC is equivalent across boys and girls with respect to (a) the pattern of factor loadings, including the common cross-loading, is equivalent, and (b) the factor variances of, and covariances among, the specified four facets of SC.7

7In contrast to the example presented in Application 1, the extent to which the error covariances are invariant across gender, within the context of this application, is of little interest. Whereas the former parameters were related to important information rooted in the interpretation of item content, the error covariance parameters in this example are related to subscale scores within the same measuring instrument.
Step 1: Testing for Validity of the Hypothesized Model Across Adolescent Boys and Girls

As with Application 1, prior to conducting any tests for invariance, we test, first, for the validity of the SC structure as represented by the multigroup model portrayed in Figure 9. In contrast to the example illustrated in Application 1, however, the multigroup model under test here comprises different hypothesized structures across groups. Turning first to specifications for adolescent boys, we see that, over and above the initially hypothesized model, there are two cross-loadings, in addition to three error covariances. By way of contrast, specifications for adolescent girls include only one cross-loading and three error covariances; these parameters replicate those for boys. Thus, prior to testing for invariance, we already know that one parameter is different across the two groups. As such, this parameter is estimated freely for boys, and is not constrained equal across boys and girls. With the exception of this parameter, then, all remaining estimated parameters can be tested for their invariance across groups. This situation serves as an excellent example of “partial measurement invariance” (see Byrne et al., 1989).

Model assessment. Goodness-of-fit statistics related to this two-group unconstrained model of SC revealed an excellent fit to the data as indicated by the CFI (.979) and RMSEA (.043) values, respectively. The chi-square value (163.362, 67 df) provides the baseline value against which all subsequent tests for invariance are compared.

TESTING FOR MULTIGROUP INVARIANCE USING AMOS GRAPHICS

As demonstrated in Application 1, prior to testing for the validity of the hypothesized multigroup model, the name of each group, together with the related data files were identified to the program. Unlike the example presented in Application 1, however, this example entails a differential number of estimated parameters for boys and girls, thereby making the testing for invariance less straightforward than was the case in the previous application, where the baseline models were identical across groups. Because Amos Graphics requires that only one model be submitted for analysis at any one time, a decision must be made regarding which baseline model (boy or girl) will be specified. Given that the baseline model for boys contains the additional parameter, this is the model of choice for reasons that will become evident as we walk through the process. Accordingly, the fully constrained (i.e., labeled) model to be tested for invariance is displayed in Figure 10.

Within the Amos Graphics framework, two possible strategies may be used in specifying a constrained multigroup model in which the baseline model differs
across groups. One strategy works from a specification perspective and calls for (in this instance) the additional cross-loading to be constrained to 0.00 for girls, albeit freely estimated for boys. The alternative strategy works from a macrolevel perspective and requests the program to allow a different path diagram for each group. We turn, first, to the specification approach.

To modify the specification of a multigroup model such that a parameter is constrained in some way for only one of the groups, we need to identify two pieces of information for the program: (a) the value to which we wish to constrain the parameter, and (b) the group for which the parameter is to be constrained. This infor-
mation is easily transmitted via the Object Properties dialog box noted in Application 1. Figure 11 shows the Object Properties dialog box that would appear after having right-clicked with the mouse on the cross-loading in question; the regression weight value has been set at 0.0. Observe, also, that the All Groups box has not been selected. Turning to the hypothesized model in Figure 12, two observations are important to note: (a) A value of zero has been assigned to the cross-loading pa-

![FIGURE 11](object_properties_dialog_box.png)

**FIGURE 11** Object Properties dialog box: Assignment of “0” to regression path. From Byrne, 2001. Reprinted with permission.

![FIGURE 12](group_names_diagram.png)

**FIGURE 12** List of group names with the female group selected: Assignment of “0” to male-specified cross-loading. From Byrne, 2001. Reprinted with permission.
rameter, and (b) In the list of groups to the left of the screen, Females has been highlighted. For clarification, it is emphasized that, in order for the zero value to be assigned to the parameter only for girls, this group must be highlighted prior to fixing the value. In contrast, if we were to highlight Males, we would see the same model, albeit with no zero value associated with the cross-loading parameter.

We turn now to the second approach to model modification whereby the program allows the groups to have different path diagrams. To initiate this strategy, we make use of the Interface Properties dialog box, which is accessed either by clicking on its icon, , or by pulling down the View/Set menu as shown in Figure 13. Several choices are available within the Interface Properties dialog box as you will readily observe in Figure 14. The one of interest to us here is the Miscellaneous option. In particular, we wish to select the boxed option listed at the bottom whereby different path diagrams are allowed for different groups. The reason why this box is not shown with a check mark in Figure 14 is because to do so automatically triggers the error message shown in Figure 15, which then blocks the information in the dialog box illustrated in Figure 14.

One peculiarity associated with the selection of this differential path diagram option is that it automatically triggers the warning message displayed in Figure 15. Although you may possibly feel a pang of anxiety about continuing after having been confronted with this caveat, you must answer “yes” if you wish to use this option. Having responded in the affirmative, the model displayed, at any point in the analysis, will be consistent with the group highlighted to its left. For example, in Figure 16, the model displayed relates to the female group; note the omission of a path leading from the ESC factor to SDQMSC. On the other hand, if we were to display the model related to the male group, we would observe this path in the diagram.

![FIGURE 13](https://example.com/fig13.png)  
**FIGURE 13** The View/Set menu. From Byrne, 2001. Reprinted with permission.
Step 2: Testing for Invariance of the Fully Constrained Model Across Adolescent Boys and Girls

Once this differential multigroup model has been established, using either of the two strategies outlined previously, testing for invariance across groups may proceed.\textsuperscript{8} Again, because it is always possible (although highly improbable) that all parameters of interest in both the measurement and structural models may be in-

\textsuperscript{8}In this example, a value of zero was assigned to the male-specific cross-loading in specifying the model for girls.
variant across boys and girls, we can test, first, for the validity of this omnibus test, which is based on the fully constrained model shown in Figure 10.9

Model assessment. A review of the goodness-of-fit indexes related to this fully constrained model once again revealed an extremely well-fitting structure, as reflected by CFI and RMSEA values of .979 and .043, respectively. The key statistic, however, is the chi-square value (224.551, 88 df), as it provides the basis for determining the extent to which the postulated model is equivalent across adolescent boys and girls. Accordingly, comparison of this model (Model 2) with the original unconstrained model (Model 1) yielded a chi-square difference value of 61.189 with 21 df, which is statistically significant (p < .001). Consistent with Application 1, analyses now proceed in a logical order that tests first for the equivalence of the measurement model and then for the equivalence of the structural model. Results bearing on this series of tests are presented in Table 2.

9For reasons cited in Footnote 7, testing for the invariance of error covariances is of little interest and, thus, these parameters remain unlabeled.
Step 3: Testing for Invariance of Factor Loadings Across Adolescent Boys and Girls

The first logical step in the invariance process is to test for the equivalence of all factor loadings across the two groups. As described and illustrated earlier, in using Amos Graphics, any reduction in the number of parameters to be tested necessarily requires that all parameter labels, except those associated with parameters under test (in this case, the factor loadings), be deleted from the model.

<table>
<thead>
<tr>
<th>Model Description</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\Delta\chi^2$</th>
<th>$\Delta df$</th>
<th>Statistical Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Combined baseline models (males and females)</td>
<td>163.36</td>
<td>67</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Factor loadings, variances, and covariances constrained equal</td>
<td>224.55</td>
<td>88</td>
<td>61.19</td>
<td>21</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>3. Factor loadings constrained equal</td>
<td>171.17</td>
<td>75</td>
<td>7.81</td>
<td>8</td>
<td>$ns$</td>
</tr>
<tr>
<td>4. Model 3 with all variances constrained equal</td>
<td>190.70</td>
<td>79</td>
<td>27.34</td>
<td>12</td>
<td>$p &lt; .01$</td>
</tr>
<tr>
<td>5. Model 3 with variance of GSC constrained equal</td>
<td>172.36</td>
<td>76</td>
<td>9.00</td>
<td>9</td>
<td>$ns$</td>
</tr>
<tr>
<td>6. Model 3 with variances of GSC and ASC constrained equal</td>
<td>190.36</td>
<td>77</td>
<td>27.00</td>
<td>10</td>
<td>$p &lt; .01$</td>
</tr>
<tr>
<td>7. Model 3 with variances of GSC and ESC constrained equal</td>
<td>174.43</td>
<td>77</td>
<td>11.07</td>
<td>10</td>
<td>$ns$</td>
</tr>
<tr>
<td>8. Model 3 with variances of GSC, ESC, and ASC constrained equal</td>
<td>175.36</td>
<td>78</td>
<td>12.00</td>
<td>11</td>
<td>$ns$</td>
</tr>
<tr>
<td>9. Model 8 with all covariances constrained equal</td>
<td>219.22</td>
<td>84</td>
<td>55.86</td>
<td>17</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>10. Model 8 with covariance between GSC and ASC constrained equal</td>
<td>180.52</td>
<td>79</td>
<td>17.16</td>
<td>12</td>
<td>$ns$</td>
</tr>
<tr>
<td>11. Model 8 with covariances between GSC/ASC and ASC/ESC constrained equal</td>
<td>180.53</td>
<td>80</td>
<td>17.17</td>
<td>13</td>
<td>$ns$</td>
</tr>
<tr>
<td>12. Model 8 with covariances between GSC/ASC, ASC/ESC, and ESC/MSC constrained equal</td>
<td>195.49</td>
<td>81</td>
<td>32.13</td>
<td>14</td>
<td>$p &lt; .01$</td>
</tr>
<tr>
<td>13. Model 8 with covariances between GSC/ASC, ASC/ESC, and GSC/ESC constrained equal</td>
<td>181.92</td>
<td>81</td>
<td>18.56</td>
<td>14</td>
<td>$ns$</td>
</tr>
<tr>
<td>14. Model 8 with covariances between GSC/ASC, ASC/ESC, GSC/ESC, and ASC/MSC constrained equal</td>
<td>216.82</td>
<td>82</td>
<td>53.46</td>
<td>15</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>15. Model 8 with covariances between GSC/ASC, ASC/ESC, GSC/ESC, and GSC/MSC constrained equal</td>
<td>182.87</td>
<td>82</td>
<td>19.51</td>
<td>15</td>
<td>$ns$</td>
</tr>
</tbody>
</table>

Note. $\Delta\chi^2 = \text{difference in chi-square values}; \Delta df = \text{difference in degrees of freedom}. \ GSC = \text{general self-concept}; \ ESC = \text{English self-concept}; \ MSC = \text{mathematical self-concept}; \ ASC = \text{academic self-concept}. \ From \ Byrne, \ 2001. \ Reprinted \ with \ permission.

aAll models compared with Model 1.
Model assessment. As indicated in Table 2, findings revealed all factor loadings to be equivalent across boys and girls, as reflected in a chi-square difference between the model tested (Model 3) and Model 1, which was not statistically significant. Given these findings, we can feel confident that all measures of SC are operating in the same way for both groups, and we proceed in testing for the equality of the structural parameters.

Step 4: Testing for Invariance of Factor Variances and Covariances Across Adolescent Boys and Girls

We test next for the invariance of factor variances across groups. Given findings of a fully invariant factor-loading matrix, model specification includes equality constraints on the factor loadings, as well as on the factor variances. Of course, as illustrated earlier, a revised model that reflects the parameters to be constrained equal (i.e., labeled parameters) must accompany each test for invariance.

In testing for the invariance of factor variances and covariances across gender, interest focuses on the hypothesized underlying factors of SC, as well as on their interrelational structure. As noted in Application 1, the testing of invariance hypotheses involves increasingly restrictive models. Thus, the model to be tested here (Model 4) is more restrictive than Model 3 because, in addition to equality constraints being imposed on the factor variances, they are also maintained for all factor loadings.

Model assessment. As detailed in Table 2, results from the estimation of Model 4 yielded a $\chi^2(79) = 190.698$. Because the difference in chi-square value between this model and Model 1, $\Delta \chi^2(12) = 27.336$, was statistically significant ($p < .01$), the hypothesis of invariant factor variances must be rejected. Faced with these results, the next task is to determine which variances are contributing to this inequality. Thus, we now proceed in testing, independently, for the invariance of each factor variance parameter while continuing to hold constrained all parameters found to be cumulatively invariant across adolescent boys and girls.

For pedagogical purposes, let us briefly review the steps involved in this testing phase. Turning to Table 2, we see that the variance of GSC was found to be invariant (Model 5). Thus, this parameter was held invariant, whereas the variance of ASC was tested for its equivalency across groups (Model 6). Results of this test revealed the latter to be noninvariant across gender; the equality constraint for the variance of the ASC factor was therefore released. As a consequence, the model used in testing for the invariance of ESC (Model 7), included two equality constraints—one for the variance of GSC, and the other for the variance of ESC. Based on this general procedure of cumulatively maintaining equality constraints only for invariant elements, all factor covariances were similarly tested. As indicated in Table 2, these tests revealed two covariances to be nonequivalent across gender.
Overall, as indicated by the goodness-of-fit statistics, the results summarized in Table 2 reveal SC structure to be well described by a four-factor model covering the facets of GSC, ASC, ESC, and MSC for both adolescent boys and girls. However, whereas the observed measures were found to be operating equivalently for both sexes, there were some differences in structural relations among the SC facets. In particular, significant gender differences were found with respect to the variance of ASC, and with respect to the covariances between ASC and MSC, and between ESC and MSC. A final model of SC structure, with all invariant parameters labeled, is displayed in Figure 17; note the absence of labels with respect to the

following parameters: variance of ASC, covariance between ESC and MSC, and covariance between ASC and MSC. (Readers are referred to Byrne, 2001, for a detailed review and explanation of the Amos Graphics output files, and to Byrne & Shavelson, 1987, for an extended discussion of the study results.)

CONCLUSION

Although, historically, issues related to the equivalency of measuring instruments and the underlying latent constructs they were designed to measure have been largely ignored in research concerned with group comparisons, the past few years have witnessed a gradual increase in the number of studies reporting findings from tests for multigroup invariance based on ANCOV structures. This recent time period has also witnessed increased interest by researchers in using the graphical interface of the Amos program (Amos Graphics), albeit the majority of applications have focused on single-group ANCOV structures. Indeed, there has been little to no evidence of its use in the conduct of tests for multigroup invariance. Based on two applications, each of which incorporated a different use of the graphical interface, this article sought to illustrate the process involved in conducting these multigroup analyses using the Amos program. It is hoped that readers interested in advancing both their knowledge of multigroup analyses and their application skills in using the Amos program will find this didactic presentation to be helpful in fulfilling their endeavors.

ACKNOWLEDGMENT


REFERENCES


