# Multilevel Latent Variable Modeling in Multiple Populations

Bengt Muthén
Siek-Toon Khoo
Jan-Eric Gustafsson *

Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90024-1521

November, 1997

*The research of the first author was supported by grant K02 AA 00230-01 from NIAAA. This research was also supported by the MARW and MALI projects in collaboration with Jan-Eric Gustafsson. We thank Ginger Nelson Goff for expert assistance.



#14

## Abstract

Modeling is described for the simultaneous analysis of two-level data in several populations. A typical example is cluster sampling of students within schools, where schools of different types are represented, e.g. public and private schools. Multivariate measurements on each student are assumed to give rise to a latent variable model. Of interest is to study across-population differences and similarities with respect to the within- and between-group covariance matrices and with respect to the mean vector. The methodology is illustrated by a comparative analysis of achievement structures in Catholic and public schools.

# 1 Introduction

Latent variable modeling of multilevel data using existing structural equation modeling software has been described in Muthén (1989, 1990, 1994a) for covariance structure models and in Muthén (1990, 1994b) for mean and covariance structure models. A typical example is cluster sampling in large-scale educational surveys with students sampled within randomly sampled schools. In many cases, the clusters (e.g. schools) have quite different characteristics and cannot be assumed to be sampled from a single common population of clusters. It is therefore of interest to generalize the mean and covariance structure modeling of multilevel data to the analysis of multiple populations. This paper develops such a multiple population model and shows how to estimate and test the model using existing structural equation modeling software.

# 2 Modeling

Consider the following data structure. Let $y_{gci}$ denote a vector of variables for a randomly sampled individual $i$ within a randomly sampled cluster $c$ for group (population) $g$. Decompose $y_{gci}$ into between- and within-cluster variation,

$$y_{gci} = y_{B_{gc}} + y_{W_{gi}} \tag{1}$$

where

$$E(y_{gci}) = \mu_{y_g} \tag{2}$$

for all $c$ and $i$ values. Consider the decomposition of the corresponding (total) covariance matrix into a within- and between-cluster part,

$$\Sigma_{T_g} = \Sigma_{B_g} + \Sigma_{W_g}. \tag{3}$$

This paper will consider latent variable models with a conventional factor analytic structure for both the between-cluster and within-cluster level. Assume for the moment that no cluster-level variables are observed. As will be described in the software implementation section, it is straightforward to include such cluster-level variables in the modeling. For the between level we specify

$$y_{B_{gc}} = \nu_g + \Lambda_{B_g} \eta_{B_{gc}} + \epsilon_{B_{gc}} \tag{4}$$

where $\nu_g$ is an intercept parameter vector, $\Lambda_B$ is a between-level loading parameter matrix, $\eta_B$ is a latent between-level variable vector, and $\epsilon_B$ is a between-level residual vector. Here,

$$E(\eta_{B_{gc}}) = \alpha_g, \tag{5}$$

$$V(\eta_{B_{gc}}) = \Psi_{B_g}, \tag{6}$$

$$V(\epsilon_{B_{gc}}) = \Theta_{B_g}. \tag{7}$$

For the within-cluster level we specify

$$y_{W_{gi}} = \Lambda_{W_g} \eta_{W_{gci}} + \epsilon_{W_{gci}} \tag{8}$$

where arrays are defined as for the between equation but where $y_{W_{gi}}$ is defined as within-cluster variation with mean zero so that the intercept vector is zero and

$$E(\eta_{W_{gci}}) = 0 \tag{9}$$

The factor analysis structure can be extended to include a structural regression structure on $\Psi_B$ and $\Psi_W$ using additional parameter arrays as in conventional structural equation modeling.

The specification of the mean structure deserves special attention in the multiple-population multilevel model. Note that we specify $E(y_{W_{gi}}) = 0$ while $E(y_{B_{gc}}) = \nu_g + \Lambda_{B_g}\, \alpha_g$ so that the means appear on the between level only. This is in line with multiple-population analysis in conventional single-level analysis. As in single-level analysis the means are specified for the level of variation for which we have independent observations, in this case the between level. It is in general not possible to also identify the within-level factor mean and it is in general not necessary to let it deviate from zero in that such across-population differences can be captured in $\nu$ and $\alpha_g$.

As an example, assume that for students sampled within schools there is a set of achievement measures which can be represented as a conventional one-factor model on the student level,

$$\Sigma_{W_g} = \lambda_W\ \psi_{W_g}\ \lambda'_W + \Theta_{W_g} \tag{10}$$

so that the factor loadings are invariant across the different school populations while the factor variance and the residual variances differ across the school populations. The difference in factor variance across school populations may correspond to different degrees of heterogeneity in student ability within school populations. Assume further that the school-level variation for these achievement variables can also be described as a one-factor model, where the factor may represent across-school variation due to selection of students into schools of different quality,

$$E(\eta_{B_{gc}}) = \nu + \lambda_B \, \alpha_g \tag{11}$$

$$\Sigma_{B_g} = \lambda_B \, \psi_{B_g} \, \lambda'_B + \Theta_{B_g} \tag{12}$$

Here we first note that $\lambda_B$ is different from $\lambda_W$ because the former reflects school-level selection and quality of instruction while the latter takes the conventional role of reflecting the measurement characteristics of the achievement variables on the student level. The across-population invariance assumption of $\lambda_W$ reflects the fact that the same measures are used in all schools. The across-population invariance assumption of $\nu$ and $\lambda_B$ is a base-line hypothesis and implies that for all school populations the same achievement variables are important in capturing the across-school variability. The result of differences across school populations in quality of student intake and instruction is in this model captured in $\alpha_g$ and $\psi_{B_g}$. A model with single factor on the between level may not be sufficient in some applications. It may be the case that schools and school populations differ with respect to more than one dimension. For example, across-school variation with respect to reading achievement may be different than that of mathematics and different school populations may show larger mean differences with respect to one type of achievement than the other. The exploration of such issues has important implications in terms of quality and equity of schooling. An illustration of this type will be analyzed in the example section using achievement data from Catholic and public schools.

# 3 Estimation

Assume random sampling of clusters independently within each of the $G$ populations. In line with McDonald and Goldstein (1989) and Muthén (1989, 1990), assume that within the sample from population $g$ there are $c = 1, 2, \ldots, C_g$ independently observed clusters with $i = 1, 2, \ldots, N_{gc}$ individual observations within cluster $c$. Let $z$ represent cluster-level variables. Arrange the data vector for which independent observations are obtained as follows for the sample from group $g$.

$$d'_{gc} = (z'_{gc}, y'_{gc1}, y'_{gc2}, \ldots, y'_{gcN_{gc}}) \tag{13}$$

where we note that the length of $d_{gv}$ varies across clusters. The mean vector and covariance matrix are,

$$\mu'_{d_{gc}} = [\mu'_{z_g}, 1'_{N_{gc}} \otimes \mu'_{y_g}] \tag{14}$$

$$\Sigma_{d_{gc}} = \begin{pmatrix} \Sigma_{zz_g} & \text{symmetric} \\ 1_{N_{gc}} \otimes \Sigma_{yz_g} & I_{N_{gc}} \otimes \Sigma_{W_g} + 1_{N_{gc}} 1'_{N_{gc}} \otimes \Sigma_{B_g} \end{pmatrix} \tag{15}$$

Assuming multivariate normality of $d_{gc}$, the ML estimator minimizes the function

$$F = \sum_{g=1}^{G} \sum_{c=1}^{C} \{\log \mid \Sigma_{d_{gc}} \mid + (d_{gc} - \mu_{d_{gc}})' \Sigma_{d_{gc}}^{-1} (d_{gc} - \mu_{d_{gc}})\} \quad (16)$$

Written in this way, the parameter arrays are potentially of large size if there are many individuals per cluster. An important simplification which makes the sizes not depend on cluster size is given as (cf. McDonald & Goldstein, 1989; Muthén, 1989, 1990)

$$F = \sum_{g=1}^{G} (\sum_{d_g}^{D_g} C_{dg} \{\ln \mid \Sigma_{b_{dg}} \mid + tr[\Sigma_{B_{dg}}^{-1} (S_{B_{dg}} + N_{dg} (\bar{\nu}_{dg} - \mu_g)(\bar{\nu}_{dg} - \mu_g)')]\}$$
$$+ (N_g - C_g)\{\ln \mid \Sigma_{W_g} \mid + tr[\Sigma_{W_g}^{-1} S_{PW_g}]\}) \quad (17)$$

where

$$\Sigma_{B_{dg}} = \begin{pmatrix} N_{dg}\Sigma_{zz_g} & \text{symmetric} \\ N_{dg}\Sigma_{yz_g} & \Sigma_{W_g} + N_{dg}\Sigma_{B_g} \end{pmatrix}$$

$$S_{B_{dg}} = N_{dg} C_{dg}^{-1} \sum_{k=1}^{C_{dg}} \begin{pmatrix} z_{dgk} - \bar{z}_{dg} \\ \bar{y}_{dgk} - \bar{y}_{dg} \end{pmatrix} [(z_{dgk} - \bar{z}_{dg})'(\bar{y}_{dgk} - \bar{y}_{dg})']$$

$$\bar{\nu}_{dg} - \mu_g = \begin{pmatrix} \bar{z}_{dg} - \mu_{z_g} \\ \bar{y}_{dg} - \mu_{y_g} \end{pmatrix}$$

$$S_{PW_g} = (N_g - C_g)^{-1} \sum_{c=1}^{C_g} \sum_{i=1}^{N_{gc}} (y_{gci} - \bar{y}_{gc})(y_{gci} - \bar{y}_{gc})'$$

Here, $D_g$ denotes the number of clusters of a distinct size in population $g$, $d_g$ is an index denoting a distinct cluster size category with cluster size $N_{dg}$, $C_{dg}$ denotes the number of clusters of that size, $S_{B_{dg}}$ denotes a between-group sample covariance matrix, and $S_{PW_g}$ is the usual pooled-within sample covariance matrix.

In the single-population case ($G = 1$) Muthén (1989, 1990) pointed out that the minimization of the ML fitting function defined by equation (17) can be carried out by conventional structural equation modeling software, apart from a slight modification due to the possibility of singular sample covariance matrices for groups with small $C_d$ values. A multiple-group analysis is carried out for $D+1$ groups, the first $D$ groups having sample size $C_d$ and the last group having sample size $N - C$. Appropriate equality constraints need to be imposed across the clusters for the elements of the parameter arrays. To obtain the correct chi-square test of model fit, a separate $H_1$ analysis needs to be done (see Muthén, 1990 for details). In the present multiple-population situation there are $D_g + 1$ groups for the $g$th population and the total number of groups is therefore much larger. To capture invariance hypotheses across populations, parameter constraints are applied across the sets of groups that represent each population.

We note in (17) that the mean structure appears in the first line and not in the second. This is in line with our model specification (4) - (9) where the means appear on the between level while the within level means are zero. Equation (17) implies that conventional structural equation modeling software should use zero sample means for the within groups with zero parameter values for these groups.

Muthén (1989, 1990) also suggested an ad hoc estimator which considered only two groups in the single-population case. With multiple populations this is generalized as

$$F' = \sum_{g=1}^{G} C_g \{\ln \mid \Sigma_{B_{\bar{c}g}} \mid + tr[\Sigma_{B_{\bar{c}g}}^{-1}(S_{B_g} + \bar{c}_g(\bar{\nu}_g - \mu_g)(\bar{\nu}_g - \mu_g)')]\}$$

$$+(N_g - C_g)\{\ln \mid \Sigma_{W_g} \mid + tr[\Sigma_{W_g}^{-1} S_{PW_g}]\} \qquad (18)$$

where the definition of the terms simplifies relative to equation (17) due to ignoring the variation in group size, dropping the $d$ subscript, and for each $g$ value using $D = 1$, $C_d = C$, and $N_d = \bar{c}$, where $\bar{c}$ is the average group size (see Muthén, 1990 for details). When data are balanced, i.e. the group size is constant for all clusters in the samples from all populations, this gives the ML estimator. Experience with the ad hoc estimator for single-population covariance structure models with unbalanced data indicates that the estimates, and also the standard errors and chi-square test of model fit, are quite close to those obtained by the true ML estimator. This observation has also been made for cases where a mean structure is added to the covariance structure, see Muthén (1994b). The ad hoc estimator, termed MUML in Muthén (1994a), will be used in the application below.

The degrees of freedom in the model testing is calculated as follows. The overall test considers a completely unrestricted model which has $G\, p$ means, and $G\, p(p+1)/2$ elements each in the between and within covariance matrices, for a total of $G\ (p{+}p(p{+}1))$ parameters. Subtracting the number of parameters in the model of interest from the number in the unrestricted model gives the degrees of freedom. Chi-square difference tests are obtained as usual for nested models.

# 4 A two-population multilevel analysis

An example is chosen which represents a common analysis problem in large scale education studies of achievement. Data are from the National Education Longitudinal Study (NELS) which is a nationally representative achievement study that tested over 20,000 students. The survey was first administered in 1988 for eighth graders with follow-up tests in the tenth and twelfth grades. In NELS, an average of about 20 students are sampled within each of a set of schools. Our analysis considers data from eighth grade with 1,044 students in 40 urban, Catholic schools and 4154 students in 195 urban, public schools. Two populations of schools are therefore considered, urban Catholic and urban public, and we view the schools as two random samples of schools from these two populations.

## 4.1 Variables and measurement models

The NELS achievement test covers reading (21 items), math (40 items), science (25 items), and history/citizenship/geography (HCG; 30 items). The variables analyzed are testlets created from these items in line with Rock et al. (1990) who considered five reading variables (literature, science, poetry, biography, history), four math variables (algebra, arithmetic, geometry, and probability), four science variables (earth, chemistry, life, methods), and three HCG variables (geography, citizenship, and history).

For the 16 achievement variables, a latent variable model was formulated for both the within-school (student) variation and the between-school variation. On the within level, a general factor ($G_W$) was specified to underlie all 16 variables with three specific residual factors (Math, Science, HCG) corresponding to math, science, and HCG variables. Holzinger and Swineford (1939) applies the same model with each variable being influenced by a general factor and a specific factor, which they called the bifactor solution. This model will be referred here as a GS model (general factor, specific factor model). Typically, the specific factors are uncorrelated among themselves and with the general factor. If covariates are included in the model, then the factors can be correlated as a function of their common dependence on the covariates. The different interpretations afforded by the higher order model where a common second-order factor influences the first-order factors and the GS model have been discussed at length in Gustafsson and Balke (1993) where the GS models are referred to as the nested factor models.

Preliminary investigations using conventional covariance structure analysis on the pooled-within matrix found it necessary to let the methods testlet of science load also on the math factor, presumably due to mathematics content in these items.

On the between level, a general factor was also specified. This general factor ($G_B$) is thought to represent school-related phenomena having to do with selection of students, neighborhood socio-economic status, and school quality. Preliminary multilevel analyses pointed to the need for including a specific, residual math factor also on the between level ($Math_B$) whereas there was no need to include specific factors for science or HCG as they do not vary significantly across schools. For each factor one loading is fixed and the remaining ones free. Figure 1 shows a path diagram of this two-level measurement model with the general and specific factors at both levels.

## 4.2 Hypotheses

It is of particular interest to study the differences across the two school populations in the means and variances of the two between factors. One may ask if the common finding of Catholic school performance advantage extends to a math performance advantage for given overall performance. The comparison of math performance without controlling for the overall performance is very much confounded by student selection and educational environment. The difference in the mean of the residual math factor may give the answer to this question. The general-specific factor measurement structure offers a way of comparing the residual specific factors conditional on the general factor, for example, the math performance can be compared conditional on the general performance. With this measurement model, the comparison of math performance across the two populations is not confounded as much by student selection issues. The presence of a math advantage may then be related to the relative emphasis that Catholic and public schools put on the math curriculum and the availability of advanced math classes in the two school systems. A secondary goal is to study if, as is assumed as a base-line hypothesis, the two school populations have the same student-level factor structure, and if the factor variances are the same across populations for these student-level factors.

A sequence of tests can be made to study invariance across school populations. This sequence is somewhat analogous to what is used for multi-population analysis in conventional, single-level analysis. In the two-level analysis, the highest level of school plays the role of the observational unit in conventional analysis. In addition to the usual choice of degree of invariance across populations for the highest observational unit, what is new in the multilevel analysis is the addition of a simultaneous choice of the within structure. First, it is of interest to see if the postulated latent variable model fits in both populations. This can be studied by a separate analysis of each of the two samples. Second, it is of interest to see if the added hypothesis of measurement invariance of between loadings and intercepts holds. In this case, the within measurement structure can be allowed to vary across the populations in terms of within loadings and within factor variances. Third, if the previous hypothesis is not rejected, it is of interest to add invariance for the within structure. Irrespective of this outcome, and if the second hypothesis is not rejected, we can study the across-population differences in the between factor means and variances.

The study of measurement invariance across populations is first carried out with the 16 achievement variables. Having established invariance at the between level, factor means and variances are compared across populations. Finally, the model is extended by including two individual covariates at the within level, gender and individual SES, and two covariates at the between level school, school-level SES and minority enrollment. The effects of these predictors of achievements are compared across populations.

## 4.3 Measurement invariance across populations

Using the MUML estimator (Muthén, 1989, 1990), a series of four-group (public-between, public-within, Catholic-between and Catholic-within) analyses were carried out to test measurement invariance across the two school populations. These are shown as model 1 to model 4 in Table 1. The approximate MUML chi-square test would reject all these four models with chi-square values ranging from 832.6 to 1139.0 with 380 to 575 degrees of freedom (p=0.000). This is to be expected with a large total sample size of 5,198.

Table 1

In model 1, all parameters are allowed to be different across populations with no assumptions of equality for the within and between factor loadings, the between y-intercepts, the factor means and variances. The chi-square value of 832.6 with 380 degrees of freedom indicates a reasonable overall fit of the measurement model in both populations given the large sample size. In model 2, the factor loadings and the y-intercepts at the between level are equated across populations to test measurement invariance at the school level. This invariance is tested by comparing model 1 and model 2. The chi-square difference is 80.6 with an increase of 35 degrees of freedom. This does not indicate a significant difference in model fit, and therefore measurement invariance at the between level is not rejected. This suggests that the 16 testlets are behaving in a similar way in both populations in capturing the across-school variability in the overall general performance and in the math performance. This invariance will enable the comparisons of the general factor performance and the between residual math factor performance across the two school populations.

Invariance of the within measurement structure is tested in model 3 by equating the factor loadings at the within level across populations. The added restrictions of model 3 relative to model 2 give 24 df and a chi-square increase of 46.0. Therefore, the invariance across populations of the within measurement structure is also not rejected. This indicates that the measurement characteristics of the testlets are not different for the public school students and the Catholic school students.

These results suggest that there is measurement invariance across populations at both the between and the within levels. The estimates for factor means and variances from the analysis specifying across-population measurement invariance on both the between and within level are shown in Table 2. The model has a chi-square value of 959.2 with 439 degrees of freedom (p=0.000) which shows acceptable fit for this large sample. We note that all factor variance components are significantly different from zero as assessed by MUML's approximate standard errors. The within structure factor variances are similar across the two school populations. On the between level, the significantly higher general factor ($G_B$) mean and the smaller variance for Catholic schools reflect a more homogeneous, better performing student population than in public schools. The residual math factor ($Math_B$) on the between level, however, has a significantly lower mean for Catholic schools. This indicates that given the same overall performance, the public schools actually perform better in mathematics. The Catholic schools general performance advantage does not extend to the performance in mathematics after controlling for overall performance.

## 4.4 Comparing effects of some predictors of achievements across populations

The above measurement model was extended to include two individual covariates, gender and individual SES, at the within level and two school covariates, minority enrollment and school-level SES, at the between level. The individual SES is a variable with four categories that reflect low to high socio-economic status of individual students while school SES is the between-level component of this variable. The minority enrollment (MinorityC) is a school variable with eight categories that reflect from low to high the proportion of minority student enrollment in the school.

Table 2

The means and standard deviations of average school SES are 2.23 and 0.59 in the public school population and 2.90 and 0.30 in the Catholic school populations. MinorityC in the public schools has a mean of 5.07 and a standard deviation of 1.76. The corresponding values are 2.50 and 2.32 in the Catholic schools. This shows that, generally, the Catholic schools are of slightly higher SES and have lower minority enrollment.

It is interesting to see how the influences of these covariates on the performances in the two school populations compare. The individual covariates were allowed to influence the $G_W$, Math, Science and HCG at the within level, and the school variables were allowed to influence the $G_B$ and $Math_B$ at the between level. This model corresponds to model 4 of Table 1, the path diagram of which is shown on Figure 2. The chi-square value is 1139.0 with 575 degrees of freedom. In the regressions of the achievement factors, the regression coefficients were allowed to be different across populations. The results are shown in Table 3. At the within level, the effect of SES is significant for the general factor in both populations. The effect on $G_W$ is positive and stronger in the public school population (b=0.24) than in the Catholic school populations (b=0.14). SES also has a positive and significant effect on the within residual math factor in the public school population (b=0.04) but a weaker and non-significant effect in the Catholic school population. This shows that, after controlling for the general performance, SES still has an effect on math performance for children in public schools but in the Catholic school population, SES does not have an effect on the residual specific factors beyond general performance. The Gender effects are strong and significant for all the within factors, general and specific, in both school populations. Results show that the girls generally do better than the boys on $G_W$, more so in the Catholic schools than in the public schools. But given the same level of general performance, the boys perform better in math, science and HCG in both school populations. The R-squares of the regressions are very similar across the two populations except for the regression of the within science factor, where the R-square is 42% for Catholic population and 23% for the public populations. R-squares for the other factors range from 7% to 15%.

Table 3

At the between level, the regression intercepts of $G_B$ and $Math_B$ for the public school population were set at zero while those for the Catholic population were estimated. The R-squares are very high compared to the within-level regressions ($G_B$: 0.76, 0.77; $Math_B$: 0.50, 0.45), this shows that the Sch SES and MinorityC variables are able to account for a large proportion of the across-school variation in the general factor and the residual math factor in both populations. The MinorityC effects on $G_B$ are significant and negative for both the public schools (b= -0.04) and the Catholic schools (b=-0.07), with a stronger effect in the Catholic schools. Using an independent t-test, the two slopes are not significantly different at the 0.05 level (t=1.4).

Fig. 3a shows a plot of $G_B$ against MinorityC at overall mean value of SES. The average general performance line for the Catholic schools is above that of the public schools across the full range of the MinorityC values. This shows that although increase in minority enrollment has a stronger negative effect on general achievement, the Catholic schools are still ahead in $G_B$ compared to the public schools. The MinorityC effect on $Math_B$ is significant and negative for the Catholic schools (b=-0.05) but not significant for the public schools (b=-0.01). The plot of $Math_B$ against MinorityC at overall mean value of SES is shown in Fig. 3b. Since the effect is negative and the public schools have a math advantage after controlling for $G_B$ and SES, the gap due to increase in minority enrollment widens. In other words, given the same general performance and school SES, public schools are ahead in math performance and this advantage is more pronounced for schools with higher minority enrollment.

School SES has positive significant effects on both the general factor $G_B$ and the residual math factor $Math_B$ in both populations. The slopes for $G_B$ are significantly different in the two populations (t=2.7). The results show that Sch SES has a greater effect in the public schools (b=0.53) and makes a greater difference to overall achievement than in the Catholic schools (b=0.26). The slopes for $Math_B$ are not different (t=0.2) in the two populations (b=0.19 for public schools and 0.22 for Catholic schools). Fig. 3c shows a plot of $G_B$ against Sch SES at the overall mean value of MinorityC. The plot shows that for higher SES schools, the public schools actually do better in the overall performance after controlling for Minority enrollment. Fig. 3d shows that the residual mathematics gap between the two school populations is not affected by Sch SES after controlling for $G_B$ and minority enrollment though the achievement level increases significantly with higher Sch SES in both school populations.

The above results regarding SES and Sch SES are quite consistent with research findings in comparing achievements in the Catholic school sector and the public school sector (Lee & Bryk, 1989; Coleman, Hoffer & Kilmore, 1982). This research usually finds that the relationship between social background and academic achievement is weaker in Catholic schools than in public schools. Lee and Bryk (1989) found that Catholic schools are more equitable not only with respect to SES but also with respect to minority enrollment. They found that a high minority concentration negatively affects the achievement in public schools more than in the Catholic schools. This is different from our finding above. We found that a high minority concentration negatively affects the general achievement in Catholic schools more than in the public schools after controlling for SES. This is also true for the residual math achievement after controlling for general achievement. This effect is negative and significant for Catholic schools but not significant for public schools.

# 5 Conclusions

This paper described a methodology for multivariate latent variable modeling of two-level data where differences and simililarities across several populations are of particular interest. This generalizes conventional latent variable multiple-group analysis to two-level data. The model imposes a structure on the between- and within-cluster covariance matrices as well as the means of the variables. Across-population comparisons of latent variable means and covariance matrices are possible when across-population invariance of between-level intercepts and loadings has not been rejected. In addition, within-level invariance of factor loadings and covariance matrices can also be tested.

To illustrate the new analysis possibilities, the methodology was applied to achievement data from Catholic and public schools. Here, data from the urban Catholic and urban public schools in the National Educational Longitudinal Study (NELS) were used. A measurement model with a general and several specific factors was formulated and tested for differences and similarities across the two populations. The within measurement structure and the between structure were allowed to be different in the two populations. Measurement invariance of the between structure across the two populations was not rejected which enabled a multidimensional comparisons of performance between the Catholic school population and the public school population. This multiple-group multilevel latent variable modeling method extends the way we can compare the relationships between social background and academic achievement across populations. For example, in our illustration, we compared influences on math performance controlling for overall performance which reduced the confounding by student selection.

# References

Bock, R. D. (1989). Multilevel analysis of educational data. San Diego, CA: Academic Press.

Bryk, A. S. and Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage Publications.

Coleman, J.S., Hoffer, T., Kilmore, S.B. (1982). High School Achievement: Public, Catholic and Other Schools Compared. New York: Basic Books.

Goldstein, H. I. (1987). Multilevel models in educational and social research. London: Oxford University Press.

Gustafsson, J. and Balke, G. (1993). General and specific abilities as predictors of school achievement. Multivariate Behavioral Research, 28 (4), 407-434.

Holzinger, K. J. and Swineford, F. (1939). A Study in Factor Analysis: The Stability of a Bi-Factor Solution. Chicago, IL: University of Chicago.

Jöreskog, K. G. and Sörbom, Dd. (1977). Statistical models and methods for analysis of longitudinal data. In D. J. Aigner and A. S. Goldberger (eds.), Latent Variables in Socio-Economic Models (pp. 285-325). Amsterdam: North-Holland.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. Biometrics, 38, 963-974.

Lee, V., and Bryk, A.S. (1989). A multilevel model of the social distribution of high school achievement. Sociology of Education, 62, 172-192.

McArdle, J. J. and Epstein, D. (1987). Latent growth curves within developmental structural equation models. Child Development, 58, 110-133.

McDonald and Goldstein (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. British Journal of Mathematical and Statistical Psychology, 42, 215-232.

Meredith, W. and Tisak, J. (1984). "Tuckerizing" curves. Paper presented at the Psychometric Society annual meetings, Santa Barbara, CA.

Meredith, W. and Tisak, J. (1990). Latent curve analysis. Psychometrika, 55, 107-122.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. Journal of Econometrics, 22, 43-65.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585.

Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting in Princeton, New Jersey. UCLA Statistics Series no. 62, August 1990.

Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins and J. Horn (eds.), Best Methods for the Analysis of Change. Recent Advances, Unanswered Questions, Future Directions (pp. 1-17). Washington D.C.: American Psychological Association.

Muthén, B. (1993). Latent variable modeling of growth with missing data and multilevel data. In C. R. Rao and C. M. Cuadras (eds.), Multivariate Analysis: Future Directions 2 (pp. 199-210). Amsterdam: North-Holland.

Muthén, B. (1994a). Multilevel covariance structure analysis. In J. J. Hox and I. G. Kreft (eds.), Multilevel analysis methods, a special issue of Sociological Methods & Research (pp. 376-391). Thousand Oaks, CA: Sage Publications.

Muthén, B. (1994b). Latent variable growth modeling with multilevel data. Presented at the UCLA conference Latent Variable Modeling with Applications to Causality. March, 1994.

Muthén, B. and Satorra, A. (1993). Complex sample data in structural equation modeling. In P. V. Marsden (Ed.), Sociological Methodology (pp.267-316). Washington D.C.: Ther American Sociological Association.

Wheaton, B., Muthén B., Alwin, D., and Summers, G. (1977). Assessing reliability and stability in panel models. In D. R. Heise (ed.), Sociological Methodology 1977 (pp. 84-136). San Francisco: Jossey-Bass.

Table 1. Multi-level multi-population analysis

---

NELS88: Public sample (195 schools, 4154 students)
Catholic sample ( 40 schools, 1044 students)

---

| | | | df | chi-sq | df-diff | chi-sq-diff |
|---|---|---|---|---|---|---|
| Model 1 | Within: | $G_W$, Math, Sc, HCG | | | | |
| | Between: | $G_B$, $Math_B$ | | | | |
| • factor loadings (within & between) not equated across pop. | | | | | | |
| • between Y intercepts not equated across pop. | | | | | | |
| • factor variances allowed to be different across pop. | | | 380 | 832.6 | | |
| Model 2 | Within: | $G_W$, Math, Sc, HCG | | | | |
| | Between: | $G_B$, $Math_B$ | | | | |
| • factor loadings (between) equated across pop. | | | | | | |
| • factor loadings (within) not equated across pop. | | | | | Model 2-1 | |
| • between Y intercepts equated across pop. | | | | | | |
| • factor variances allowed to be different across pop. | | | 415 | 913.2 | 35 | 80.6 |
| Model 3 | Within: | $G_W$, Math, Sc, HCG | | | | |
| | Between: | $G_B$, $Math_B$ | | | | |
| • factor loadings (within & between) equated across pop. | | | | | Model 3-2 | |
| • between Y intercepts equated across pop. | | | | | | |
| • factor variances allowed to be different across pop. | | | 439 | 959.2 | 24 | 46.0 |
| Model 4 | Within: | $G_W$, Math, Sc, HCG | | | | |
| | Between: | $G_B$, $Math_B$ | | | | |
| **(With 2 school-level covariates, MinorityC and Sch SES, and 2 individual covariates, SES and Gender)** | | | | | | |
| • factor loadings (within & between) equated across pop. | | | | | | |
| • between Y intercepts equated across pop. | | | | | | |
| • factor variances allowed to be different across pop. | | | | | | |
| • Z means and reg. coeffs. not equated across pop. | | | 575 | 1139.0 | | |

---

Table 2: Model with measurement invariance on both the between and within levels: no covariates

$\chi^2(439) = 959.2, \quad p=0.000$

| | Public Urban n=4154 | | Catholic Urban n=1044 | |
|---|---|---|---|---|
| **Within** | | | | |
| Factor variances | | | | |
| $G_W$ | 0.58 | (0.02)† | 0.52 | (0.03) |
| Math | 0.26 | (0.02) | 0.26 | (0.03) |
| Science | 0.16 | (0.05) | 0.10 | (0.05) |
| HCG | 0.11 | (0.04) | 0.19 | (0.04) |
| **Between** | | | | |
| Factor variances | | | | |
| $G_B$ | 0.27 | (0.04) | 0.12 | (0.04) |
| $Math_B$ | 0.02 | (0.01) | 0.06 | (0.02) |
| Factor means | | | | |
| $G_B$ | 0.00 | | 0.57 | (0.08) |
| $Math_B$ | 0.00 | | -0.18 | (0.05) |

† Standard errors are given in parentheses

Table 3: Model with measurement invariance on both the between and within levels: covariates included

| $\chi^2(575) = 1139.0$, p=0.000 | | | | |
|---|---|---|---|---|
| | Public Urban n=4154 (195 schools) | | Catholic Urban n=1044 (40 schools) | |
| Residual Variances | | | | |
| **Within** | | | | |
| $G_W$ | 0.52 | (0.02)† | 0.49 | (0.03) |
| Math | 0.26 | (0.02) | 0.26 | (0.03) |
| Science | 0.17 | (0.03) | 0.11 | (0.04) |
| HCG | 0.04 | (0.01) | 0.03 | (0.01) |
| **Between** | | | | |
| $G_B$ | 0.04 | (0.01) | 0.02 | (0.01) |
| $Math_B$ | 0.02 | (0.01) | 0.04 | (0.02) |
| Intercepts | | | | |
| **Between** | | | | |
| $G_B$ | 0.00 | | 0.83 | (0.26) |
| $Math_B$ | 0.00 | | -0.17 | (0.29) |
| Regression coefficients | | | | |
| **Within** | | | | |
| Regression of $G_W$: | $R^2$= 0.09 | | $R^2$= 0.07 | |
| SES | 0.24 | (0.02) | 0.14 | (0.03) |
| Male | -0.19 | (0.03) | -0.30 | (0.05) |
| Regression of Math: | $R^2$= 0.07 | | $R^2$= 0.10 | |
| SES | 0.04 | (0.01) | 0.04 | (0.03) |
| Male | 0.26 | (0.03) | 0.35 | (0.05) |
| Regression of Science: | $R^2$= 0.23 | | $R^2$= 0.42 | |
| SES | 0.01 | (0.02) | 0.02 | (0.03) |
| Male | 0.45 | (0.03) | 0.57 | (0.06) |
| Regression of HCG: | $R^2$= 0.13 | | $R^2$= 0.15 | |
| SES | 0.00 | (0.01) | -0.01 | (0.01) |
| Male | 0.16 | (0.02) | 0.16 | (0.03) |
| **Between** | | | | |
| Regression of $G_B$: | $R^2$= 0.76 | | $R^2$= 0.77 | |
| Minority | -0.07 | (0.02) | -0.07 | (0.02) |
| School SES | 0.53 | (0.07) | 0.26 | (0.06) |
| Regression of $Math_B$: | $R^2$= 0.50 | | $R^2$= 0.45 | |
| Minority | -0.01 | (0.01) | -0.05 | (0.02) |
| School SES | 0.19 | (0.07) | 0.22 | (0.08) |

† Standard errors are given in parentheses

Table 3 contd.

Factor loadings

**Within**

| | $G_W$ | Math | Science | HCG |
|---|---|---|---|---|
| $R_1$ | .98 | .00 | .00 | .00 |
| $R_2$ | 1.34 | .00 | .00 | .00 |
| $R_3$ | 1.12 | .00 | .00 | .00 |
| $R_4$ | 1.28 | .00 | .00 | .00 |
| $R_5$ | 1.22 | .00 | .00 | .00 |
| $M_1$ | .98 | .86 | .00 | .00 |
| $M_2$ | 1.00 | 1.00 | .00 | .00 |
| $M_3$ | .66 | .62 | .00 | .00 |
| $M_4$ | 1.03 | .88 | .00 | .00 |
| $S_1$ | .89 | .00 | 1.00 | .00 |
| $S_2$ | .72 | .30 | .50 | .00 |
| $S_3$ | .82 | .00 | .29 | .00 |
| $S_4$ | .85 | .00 | .25 | .00 |
| $H_1$ | 1.04 | .00 | .00 | .67 |
| $H_2$ | .77 | .00 | .00 | 1.00 |
| $H_3$ | .95 | .00 | .00 | 2.18 |

**Between**

| | $G_B$ | $Math_B$ |
|---|---|---|
| $R_{1B}$ | .96 | .00 |
| $R_{2B}$ | 1.29 | .00 |
| $R_{3B}$ | 1.12 | .00 |
| $R_{4B}$ | 1.19 | .00 |
| $R_{5B}$ | 1.46 | .00 |
| $M_{1B}$ | .92 | .95 |
| $M_{2B}$ | 1.00 | 1.00 |
| $M_{3B}$ | .55 | .67 |
| $M_{4B}$ | 1.26 | .41 |
| $S_{1B}$ | 1.06 | .00 |
| $S_{2B}$ | .85 | .38 |
| $S_{3B}$ | .88 | .00 |
| $S_{4B}$ | 1.07 | .00 |
| $H_{1B}$ | 1.05 | .00 |
| $H_{2B}$ | .86 | .00 |
| $H_{3B}$ | 1.02 | .00 |

Figure 1. A two-level measurement model with general and specific factors

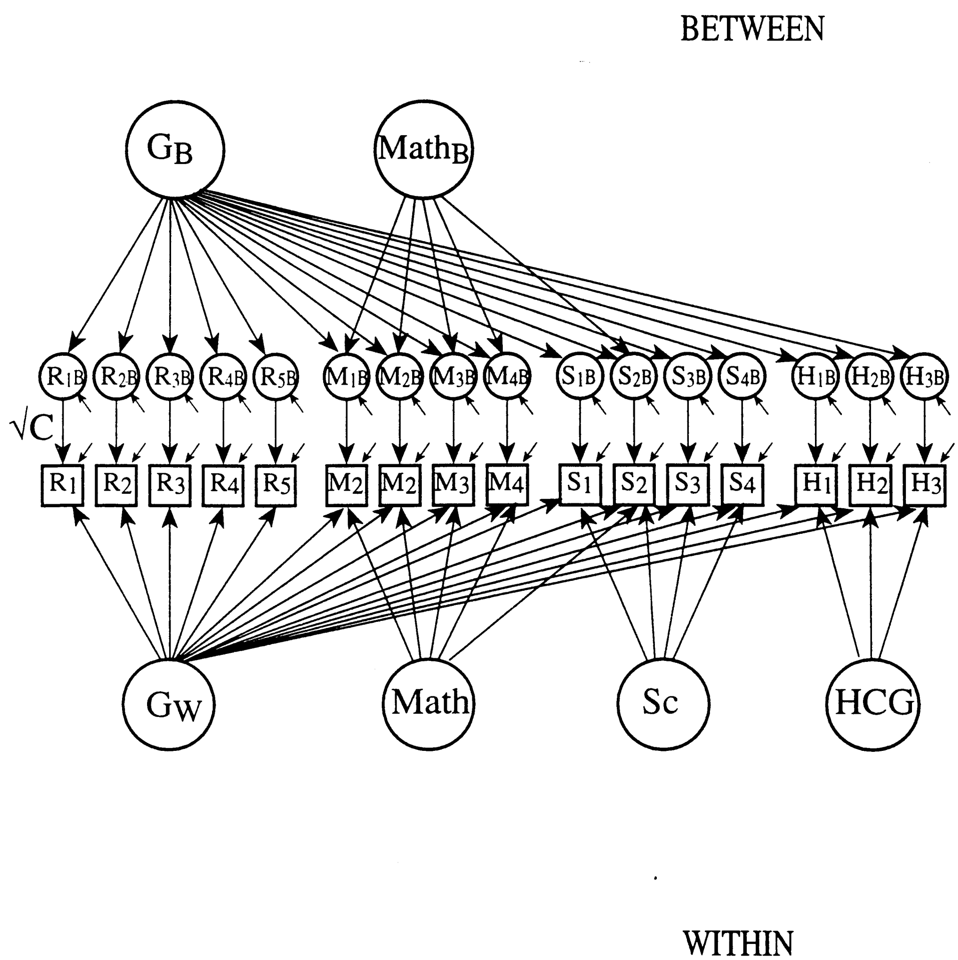

Figure 2. The two-level model with covariates

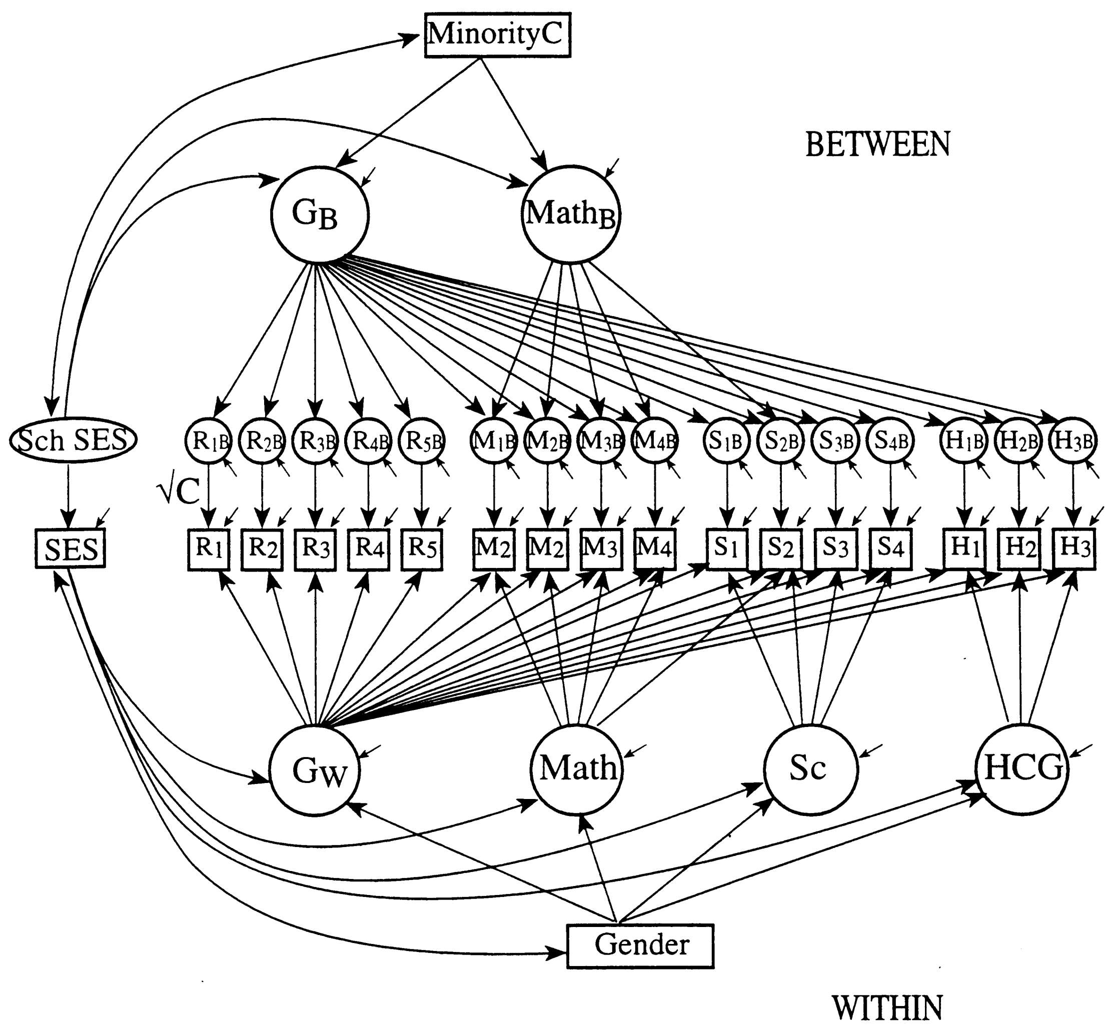

Fig 3a. $G_B$ x MinorityC (at mean SES)

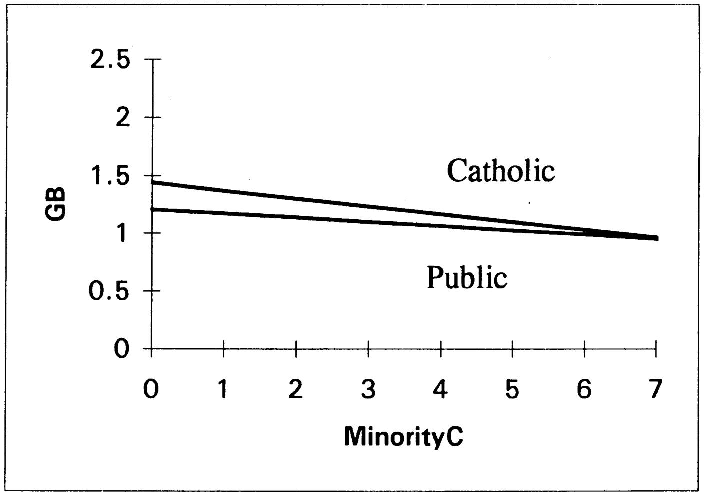


Fig 3b. $Math_B$ x MinorityC (at mean SES)

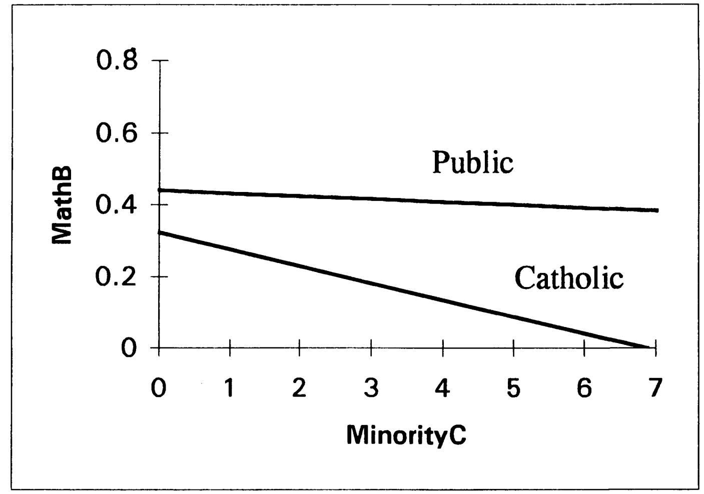


Fig 3c. $G_B$ x Sch SES (at mean MinorityC)

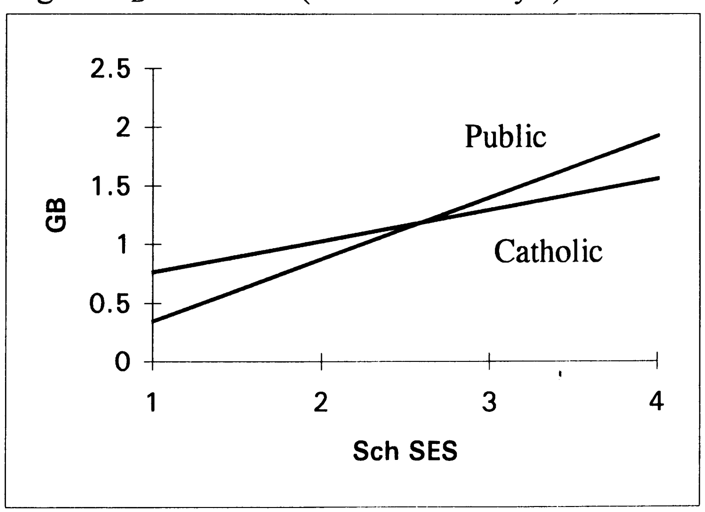


Fig 3d. $Math_B$ x Sch SES (at mean MinorityC)