This is to certify that the

dissertation entitled

Estimating the Covariance Components of an Unbalanced Multivariate Latent Random Model Via the EM Algorithm

presented by

Leonard Joseph Bianchi

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Counseling, Educational Psychology & Special Education (Statistics & Research Design)

Wm H Schmidt
Major professor

Date May 16, 1987

REMOTE STORAGE RSF

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
	AUG 22 2018	

ESTIMATING THE COVARIANCE COMPONENTS OF AN UNBALANCED MULTIVARIATE LATENT RANDOM MODEL VIA THE EM ALGORITHM

by

Leonard Joseph Bianchi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational
Psychology and Special Education

1987

ABSTRACT

ESTIMATING THE COVARIANCE COMPONENTS OF AN UNBALANCED MULTIVARIATE LATENT RANDOM MODEL VIA THE EM ALGORITHM

By

Leonard Joseph Bianchi

Although statistical procedures are available for estimating treatment effects for students taught in classrooms, these procedures are applicable only when every class has the same number of students. The present study investigated a procedure that was originally established to handle missing data (EM Algorithm) but which also provides a solution to the problem of estimating parameters in multivariate analysis when samples contain unequal group sizes. The focus of the present dissertation was on the estimation of latent group and individual level variances and covariances with measurement error removed when group sizes varied in a sample. Previous methods could only find maximum likelihood estimates for this problem if the dataset contained groups of equal size. The EM Algorithm offers a method for finding maximum likelihood estimates of parameters in situations where classical maximum likelihood procedures fail.

The estimate of balanced and unbalanced samples were both studied while varying two factors, mainly the number of groups in the sample

(the size) and the particular model being estimated (that is to say, the unrestricted model, the correctly specified model and the incorrectly specified model). Only 10 replications were used in this demonstration of the algorithm under different circumstances.

Tests of the model based on the criteria of convergence showed this estimation procedure to be a satisfactory and effective method in theory. However, problems in the use of this algorithm appearred in the form of large number of iterations needed for convergence and lack of a universally accepted criterian for convergence.

This dissertation is dedicated to Robert Garden, Chris Mooney and the small group of IEA workers in New Zealand from 1981 to 1983 from whom I learned so much.

ACKNOWLEDGMENTS

I can never fully express the appreciation and thanks I have for Dr. William Schmidt who as both a friend and my chairman never gave up on me. His support was invaluable.

To Dr. Richard Houang I owe much for both his time and effort. He forced me to think, a pesky thing to do, and attempted to open my eyes and mind and look beyond the obvious meaning of the research.

The other two members of my committee also have my thanks. Dr. Stephen Raudenbush gave freely of both his time and knowledge while Dr. Robert Floden contributed suggestions and ideas.

A very special thanks is due to JoAnn Green who, as a friend and university employee, helped steer me through the sea of regulations and red tape which, by necessity, exist in a university. Without her kindness and help, I know all would have been lost.

Thanks is also given to Janice Benjamin for the use of her equipment and to Laurence Bates who helped me with the logistics which must be faced by every doctorial student.

My gratitude goes to Dr. Arthur Tabachneck for his long hours spent in an attempt to teach me the lost art of technical writing. His kindness is more than appreciated.

My special appreciation goes to Peri who put up with my tirades, depressions and periods of paranoia with patience and care.

Lastly my appreciation goes to my parents who never faltered in their belief in me, who helped me out during the tough times and who never once asked "When are you going to graduate and get a job?". (Well, hardly ever .....)

TABLE OF CONTENTS

LIST OF TABLES vii
LIST OF FIGURES viii

Chapter

I. INTRODUCTION 1

II. LATENT COVARIANCE STRUCTURAL MODELS 6

1. Single Level Covariance Structural Model 6
2. Multilevel Covariance Structural Model 17
3. Extension of Φ_a and Φ to causal models 22

III. REVIEW OF RELEVENT WORK 23

1. Schmidt's Structural Model 23
2. Unbalanced Designs 32

IV. STATEMENT OF PROBLEM 35

V. ESTIMATION PROCEDURE 37

1. Expectation-Maximization (EM) Algorithm 37
2. Theory for the Restricted Model 39
3. The EM algorithm 48

VI. DESIGN OF STUDY 49

1. Design 49
2. Generation of Data 52
3. Starting Values of Parameters 55
4. Computer Program 57

VII. RESULTS OF STUDY 59

1. Design and Measures 59
2. Results of Estimation Procedure 62
3. Results of the Process 86

VIII. DISCUSSION 90

1. Summary and Conclusions 90
2. Future Explorations 99

APPENDICES

A. Equations for the Estimation of the Covariance Components of the Two-Parameter Model Using the EM Algorithm 101
B. Equations for the Estimation of the Covariance Components of the Four-Parameter Model Using the EM Algorithm 103
C. Computer Program for the Estimation of the Three Parameter Latent Model through the SAS Package 106
D. Computer Program for the Estimation of the Four Parameter Latent Model through the SAS Package 115

BIBLIOGRAPHY 123

LIST OF TABLES

Table		Page
1	Parameter Values Used in Data Generation	53
2	Summary Statistics of the Latent Models for Balanced and Unbalanced Data Sets with 100 Groups	63
3	Summary Statistics of the Latent Models for Balanced and Unbalanced Data Sets with 25 Groups	68
4	Summary Statistics of the Unrestricted Model for Balanced and Unbalanced Data Sets for both 25 and 100 Groups	72
5	Average B/E of the Latent Models for Balanced and Unbalanced data Sets for both 25 and 100 Groups	77
6	Average Ratio(1) of the Ph, Om and Ps Matrices for the Latent Models in Data Sets for both 25 and 100 Groups	78
7	Average Ratio(2) of the Ph, Om and Ps Matrices for Balanced and Unbalanced Data Sets for both 25 and 100 Groups	80
8	Maximum Likelihood Ratio Test of the Fit of the Correct and Incorrect Models	82
9	Average B/E of the Unrestricted Model for Balanced and Unbalanced Data Sets both 25 and 100 Groups	84
10	Average Ratio(1) of the Phi and Psi Matrices of the Unrestricted Model for Balanced and Unbalanced Data Sets for both 25 and 100 groups	85
11	Iterations Required by Algorithm for Convergence	87
12	Summary Statistics of the Four Parameter Latent Model for Balanced and Unbalanced Data Sets with 100 Groups	94

LIST OF FIGURES

Figure Page

1 Design of Study 51

CHAPTER I: INTRODUCTION

Although statistical procedures are available for estimating treatment effects for students taught in classroom groups, these procedures are only applicable when every class has the same number of students. Since equal class size is rare in schools, however, it is important to develop practical methods that extend current procedures to cover all patterns of class size. The present study investigates a procedure that was originally established to handle missing data (the EM algorithm), but which also provides a solution to the problem of unequal sample sizes in multivariate analyses. After presenting a review of existing procedures, this dissertation will: (i) show how the EM algorithm can be applied to this case; (ii) exhibit a computer program that uses this procedure to analyze such data; and (iii) illustrate the procedure and program with an analysis that estimates parameters from a sample data set generated from a known distribution.

Many educational researchers have engaged in attempting to identify the various factors which affect student achievement. Laboratory studies have identified how individuals respond to different educational treatments, but most formal education occurs in classroom settings in which students receive treatments as a group. Two effects are introduced in the latter situation, however, which cannot be overlooked.

First, there may be some process which affects the class as a whole. A teacher with a class having a mean IQ of 120 may decide to cover more material than one with a class having a mean of 100. Such an

effect, unless accounted for, could obscure the relationship between IQ and achievement.

Second, similarly, there may be an interactive effect between individuals and the class. A person with an IQ of 110 may do much differently in a class with a mean of 100 than in a class with a mean of 120. That is, class level processes can affect between class analyses, interactive processes can influence within class analyses and individuals can have an effect on both. The analysis of such data must be interpreted carefully.

Estimates of relationships between variables at the class and individual levels can either be low, or high, as two effects can either combine to indicate a spuriously high relationship or work against each other to reduce it. A number of models and strategies have been recently developed to analyze this type of data. One line was the development of regression models to study the individual and group effects. Others developed models to estimate underlying latent variances at each level.

Keesling and Wiley (1974) used the relationship between two student level variables to adjust class level scores. The aggregated values of the variables were used with the estimated student level regression coefficients to compute an expected group score. This score was then subtracted from the aggregated class scores in order to obtain residual scores which, in turn, were then used within regular linear regression models with class level variables. Keesling and Wiley's model was based on the assumption that the relationship between two variables would be identical for all levels of the model.

Cronbach (1976) proposed an analytic approach that focussed on processes going on both between and within groups. He felt regression effects were composed of two components, a between and a within effect. His model allowed for the relationship between two variables to vary between the individual and classroom levels.

Burstein, Linn and Capell (1980) developed a model allowing regression coefficients between two variables to change from class to class. These coefficients were then used as dependent variables in regression analyses at the group level. Raudenbush (1986) applied empirical Bayes theory to develop a procedure for producing Maximum Likelihood (ML) estimates of the regression coefficients in Burstein's et al model.

None of the univariate regression models, however, offered methods for estimating measurement error. Tests and instruments used in education, virtually by definition, contain measurement errors which can inflate the analyses' error term and weaken the fit of the model.

Schmidt (1971) addressed this problem by developing a multivariate structural model. By fitting an a priori structure to the covariance matrix of the student's tests, the variance and covariance of the latent dimensions and measurement errors could be estimated for both levels. This was especially significant as both the variance and covariance of latent dimensions are frequently the items of importance to researchers.

An example of the applicability of this notion was evidenced in the International Association for the Study of Educational Achievement's (IEA) "Second International Mathematics Study", where items within academic tests were systematically constructed from a number of

dimensions. In one case, one dimension was word problems vs numerical examples while another was arithmetic vs algebra. All items containing the two dimensions, in turn, could be combined to make four subtests. The subtests, theoretically, were assumed to have served as congeneric tests (see Chapter 8, Lord and Novich, 1976).

The four subtests contained the following dimensions:

	Type of Problem	Type of Mathematics
Subscore A	Word Problems	Arithmetic
Subscore B	Word Problems	Algebra
Subscore C	Numerical	Arithmetic
Subscore D	Numerical	Algebra

Of interest to the IEA study was not the four subscores but the variance and covariance of the two underlying dimensions. Schmidt's model could have been used to provide estimates of the latent variance and covariance at both the class and individual levels. These covariance matrices can then be used in further analysis based on numerous models.

Wisenbaker (1981), following the same logic, further developed the estimation procedures necessary to estimate parameters of a causal model for latent covariance structures. The structural parameters of the between and within levels, according to Wisenbaker's model, are simultaneously estimated yielding ML estimators.

Schmidt's and Wisenbaker's models, however, both require groups (classes) to be of equal size. The underlying multivariate normal

distribution upon which the ML equations are based is a vector of length np where n is the number of students in each class and p is the number of observed measures (tests) taken by each student. The ML procedure requires that each group contains the same number of subjects - n. In educational research, however, the number of subjects usually varies from classroom to classroom.

The present dissertation concentrates on the problem of estimating the latent between and within covariance matrices when the number of subjects (students) varies between groups (classes). The early chapters contain information on the background of the problem. Chapter Two, Latent Structural Models, describes the development and background of those models. Chapter Three describes the background of the specific model used in this study and the development of different techniques for estimating variance components in the unbalanced random model. Chapter Four contains a statement of the problem.

The last chapters contain the derivation of the procedure and an example of its use. Chapter Five contains the derivation of the equations needed in the estimation procedure. Chapter Six details the design of a monte carlo study for illustrating the use of the EM algorithm under the current model. The results of the study are described in Chapter Seven. Chapter Eight, finally, presents a discussion of the results and conclusions.

CHAPTER II: LATENT COVARIANCE STRUCTURE MODELS

1. Single Level Covariance Structure Model

Latent Covariance structure analysis was developed along two different lines of inquiry. The first approach, factor analysis, was derived explicitly for the purpose of finding latent structures (Spearman, 1904). The second line of inquiry applied the existing random analysis of variance model toward solving the same problem (e.g. Bock, 1960).

Factor analysis was developed by Spearman as a method for confirming his theory on ability. Spearman sought to show that IQ tests measured two components, a "general or G factor" common to all IQ tests and a second factor specific only to the test. Through the application of factor analysis, he was able to isolate the variance component of each test attributable to the G factor, as well as the variance component specific to the individual tests. As the mathematics for factor analysis were expanded and refined, however, its use changed from confirmatory to exploratory and became a method for reducing a set of items or measures to a lesser number of underlying latent dimensions. These dimensions, in turn, were used to form factor scores for discriminating between subjects. These later exploratory methods of factor analysis lacked a firm theoretical basis and were simply algebraic manipulations of the data.

A confirmatory approach to factor analysis did not resurface until the 1950's when such an approach was considered by Howe (1955), Anderson and Rubin (1956), and Lawley (1958). The advantage of

confirmatory approaches lie in their use of statistical theory and ability to test the fit of the latent models, but early efforts went largely ignored because of computational difficulties. Interest gradually rose after Joreskog (1966) developed more efficient estimation procedures.

Joreskog (1969) developed a general approach to confirmatory maximum likelihood factor analysis. Unlike the prior models, this model had the flexibility of allowing researchers the capability of selecting different structures from possible solutions: orthogonal, oblique and various mixtures of the two. The factor analysis model is based on the fundamental equation

$$\underline{y} = \underline{\mu} + \Lambda\underline{x} + \underline{z} \tag{2.1a}$$

where $\underline{y}$ is a p x 1 vector of observed variables, $\underline{\mu}$ is a vector of grand means, Λ is p x q matrix connecting the p observed values to the q latent factors (with $q \leq p$), $\underline{x}$ is a q x 1 vector of the latent factors, and $\underline{z}$ is a p x 1 vector of the error or unique parts of the test. It is assumed that $E(\underline{x})=E(\underline{z})=\underline{0}$, $E(xx')=\Phi$, $E(zz')=\Psi$, and $E(yy')=\Sigma_y$. The dispersion matrix for $\underline{y}$ is

$$\Sigma_y = \Lambda \Phi \Lambda' + \Psi$$

Assuming $\underline{y}$ has a multivariate normal distribution, the maximum likelihood equation is

$$L = \Pi_{i=1}^{n}(2\pi)^{-p/2} \left|\Sigma_y\right|^{-1/2} \exp\{ -\tfrac{1}{2} (\underline{y}_i - \underline{\mu}_y)' \Sigma_y^{-1} (\underline{y}_i - \underline{\mu}_y) \} \tag{2.3a}$$

The efficient part of the log(L) is

$$\log(L) = -\frac{1}{2} n \{ \log|\Sigma| + tr(S\Sigma^{-1}) \} \quad . \tag{2.4a}$$

Minimizing the following function

$$F(\Lambda, \Phi, \Psi) = \log|\Sigma| + tr(S\Sigma^{-1}) - \log|S| - p \tag{2.5a}$$

yields the likelihood ratio test statistic of goodness of fit.

A second approach was developed through the use of the random analysis of variance model. Burt (1947) was the first to point out the analogy between the analysis of variance (ANOVA) and factor analysis. This was further elaborated by Creasy (1954). Bock (1960) showed that a formal relationship exists between the two approaches. This relationship only becomes clear if a distinction is made between factor analysis used as a "structural" versus "discriminal" analysis. According to Bock (1960, p153):

"By 'structural' analysis is meant a measure which attempts to make causal statements about test performances by assigning to definite sources the covariation which arises between certain psychological tests: this was the original use of factor analysis. In its subsequent application to the construction of test batteries, factor analysis was also used to assess whether tests of known measurement error yield reliable distinctions between individuals, and, if so, in how many dimensions: it seems appropriate to designate this 'discriminal' analysis." Factor analysis doesn't separate these two uses or give clear answers for either.

Bock showed that a Model II (Random) ANOVA model can be applied to tests in light of specific hypothesis about their composition and suitably adjusting their psychometric characteristics. The analysis could be used to study structural and discriminal properties of the tests, free of difficult statistical and interpretation problems. The purposes of this dissertation deal only with the structural analysis and shall concentrate on it using an example to facilitate the discussion.

Consider the design of four tests from two dichotomous dimensions, as referred to in Chapter 1, namely Type of Problem ((1) Word Problems vs (2) Numerical) and Type of Mathematics ((1) Arithmetic vs (2) Algebra).

In our example, the four tests may be identified by the following ordered pairs:

B(k) \ A(j)	1	2
1	11	12
2	21	22

Test(jk)	Type of Problem	Type of Mathematics
Test 11	Word Problems	Arithmetic
Test 12	Word Problems	Algebra
Test 21	Numerical	Arithmetic
Test 22	Numerical	Algebra

A model for the structural analysis of this design is

$$X_{ijkt} = \alpha_i + \beta_{ij} + \gamma_{ik} + \delta_{ijk} + \epsilon_{ijkt} \qquad (2.6a)$$

where X_{ijkt} is the score of individual i on test jk on occasion t, α_i is a component of score specific to individual i on all tests, β_{ij} and γ_{ik} are components of score specific to individual i on dimensions B and C respectively, δ_{ijk} is a component of score specific to individual i and the test jk (with the dimensions effect excluded) and ϵ_{ijkt} is a replication error specific to individual, test, and occasion. These components are considered random effects and are assumed normal and independent,

$$\alpha \sim N(0, \sigma_a^2)$$
$$\beta \sim N(0, \sigma_b^2)$$
$$\gamma \sim N(0, \sigma_c^2)$$
$$\delta \sim N(0, \sigma_d^2)$$
$$\epsilon \sim N(0, \sigma_e^2)$$

Because the number of components with a distribution over individuals is equal to the number of tests in the dichotomous factorial design, the covariance structure may be fully estimated. This will not be true when there are more than 2 levels to a dimension.

The design for our example can be represented in matrix form as the Hamadand design matrix

(2.7a) $P = 1/2$

μ	B	C
1	1	1
1	1	-1
1	-1	1
1	-1	-1

The purpose of the structural analysis is to test whether the sample covariances between tests fit the model. The covariance matrix of the data from our example is

(2.8a) $$S_r = X.'X. - (M /(n-1))$$

where X is the N x 4 matrix of means of r replicate scores and M is the matrix of corrections to the sample data. Pre and post multiplying the population covariance matrix by P will reduce it to its cononical or diagonal form ($P \Sigma_r P'$). If the sample covariance matrix is treated in this way, the off diagonal elements will not necessarily be equal to zero but if the model fits though, any non zero value will attributed simply to sampling variance. Therefore a statistical test of the off diagonals being equal to zero will be a test of the fit of the model.

A maximum likelihood ratio test given by Wilk's criterion and a chi-squared approximation provided for moderate to large samples by Bartlett can be used to test that hypothesis (Anderson,1984). This is

(2.9a) $$X^2 = - (N - (2p + 11)/6) \log|R_r|$$

where p is the number of variates, $|R_r|$ is the determinate of ARA' and R is the correlation matrix corresponding to S.

Bock's work on reformulating factor analysis in the form of a random model has spurred the development of more complicated models and situations. Bock and Bargmann (1966) presented a method for analyzing a sample covariance matrix in order to assess the latent sources of variance and covariance within multivariate normal data. This "structural" analysis of the sample covariance matrix has a two fold purpose. The first purpose is to statistically test the feasibility of a hypothesized model and the second is to provide estimates of variance components associated with the latent variables of this model.

This analysis is an alternative to Type III (Mixed) ANOVA Model. The method of maximum likelihood estimation is used to test the model and estimate the latent variance-covariance components. The model for the observed score vector of p tests is given as

(2.10a) $$\underline{y}_{it} = \underline{\mu} + A\,\underline{\xi}_i + \underline{\epsilon}_{it}$$

where $\underline{\mu}$ is the vector mean of the p tests, A is a p x m matrix of known coefficients connecting the observed and the latent variables, $\underline{\xi}_i$ is an m x 1 vector of latent scores for subject i having an m x m covariance matrix Φ and $\underline{\epsilon}_{it}$ is a p x 1 vector of measurement errors with a p x p covariance matrix Ψ.

The model implies that vector $\underline{y}_{it}$ has a multivariate normal distribution with mean vector $\underline{\mu}$ and covariance matrix Σ_y where

(2.11a) $$\Sigma_y = A\,\Phi\,A' + \Psi$$

In this model, the latent variables are considered independent of each other with Φ considered to be a diagonal matrix.

The likelihood function of the general model proposed by Bock and Bargmann for p measures on N individuals sampled randomly from a multivariate normal population is

(2.12a) $$L = \Pi_{i=1}^{n}(2\pi)^{-p/2}\,|\Sigma_y|^{-1/2}\exp\{-\tfrac{1}{2}(\underline{y}_i - \underline{\mu}_y)'\,\Sigma_y^{-1}\,(\underline{y}_i - \underline{\mu}_y)\}$$

Taking the natural log of the function, differentiating and setting the derivitive equal to zero will yield a maximum likelihood estimate

(2.13a) $$\text{Log}(L) = -(Np/2)\log 2\pi - (N/2)\log|\Sigma_y| - (N/2)\,\text{tr}\{\Sigma^{-1}S\}$$

Assuming the elements of Σ_y are functions of a scalar variable x, the first derivitive with respect to x is

(2.14a) $$\frac{\delta \text{Log}(L)}{\delta x} = (N/2)\,\text{tr}\left\{\frac{\delta \Sigma}{\delta x}\left(\Sigma^{-1}S\,\Sigma^{-1} - \Sigma^{-1}\right)\right\}$$

The second derivitive with respect to scalars x and y is

(2.15a) $$\frac{\delta \text{Log}(L)}{\delta x\,\delta y} = \frac{N}{2}\,\text{tr}\left\{\Sigma^{-1}\frac{\delta \Sigma}{\delta y}\,\Sigma^{-1}\frac{\delta \Sigma}{\delta x}\right\} - \frac{N}{2}\,\text{tr}\left\{\Sigma^{-1}\frac{\delta^2 \Sigma}{\delta x \delta y}\right\} + \frac{N}{2}\,\text{tr}\left\{\frac{\delta W}{\delta y}\frac{\delta \Sigma}{\delta x}\right\} + \frac{N}{2}\,\text{tr}\left\{W\frac{\delta^2 \Sigma}{\delta x \delta y}\right\}$$

where $W = \Sigma^{-1}S\,\Sigma^{-1}$.

Because the scalars are not directly estimable, the Newton Raphson algorithm was used. This algorithm required the first and second derivitives of the Log likelihood and may unfortunately yield variance components with negative values.

The likelihood equations and computational scheme above has been worked out for three structural models. They were three distinct cases for the model

(2.16a) $$\Sigma = A\,\Phi\,A' + \Psi$$

Case I. The latent variables are uncorrelated, A is completely specified and unscaled, and the error variances are assumed homogeneous ($\Psi = \sigma^2 I$).

Case II. The latent variables are uncorrelated, A is completely specified and unscaled, and the error variances are assumed

heterogeneous ($\Psi = \text{diag}\,[\Psi_{11}, \Psi_{12} \ldots, \Psi_{pp}]$).

Case III. The latent variables are assumed uncorrelated, A is completely specified, but scaled by an unknown, but estimable, matrix of scaling factors (Γ), and the error variances are assumed homogeneous ($\Psi = \sigma^2 I$).

Bock and Bargmann chose the likelihood statistic

$$\lambda = - \frac{|S|^{\frac{N}{2}}}{|\hat{\Sigma}|^{\frac{N}{2}}} \tag{2.17a}$$

to test the hypothesis that the population covariance matrix has a Case I, II or III structure vs an unrestricted structure. The distribution of S in large samples may be approximated by a chi square.

$$X^2 = -2 \log\lambda = N \log(|\hat{\Sigma}|/|S|) \tag{2.18a}$$

The degrees of freedom for this statistic is equal to the difference in the number of parameters in the restricted and unrestricted models.

These three cases are only a small number of the many possible covariance structure models that can be hypothesized for the general model (2.6a).

Wiley (1967) developed a set of 16 models that can be hypothesized by applying different combinations of restrictions to the three main components of the general model. Studying other researcher's comments about the model, he allowed the latent variables to also be correlated and some of the elements of Λ to be specified a priori. The constraints he proposed for the parameter matrices of the model formed 16 possible

models (4 x 2 x 2). The Λ matrix could be constrained four different ways while Φ and Ψ could have one of two different shapes.

The four forms of restriction on Λ are

(1) General (Λ) - all elements are to be estimated.

(2) General (Λ) - most elements are to estimated except for certain specified ones

If Λ is reparameterized into ΓA where Γ is a matrix of scaling factors, two more sets of restriction can be applied.

(3) Completely specified (A) and scaled by an unknown but estimable matrix of scaling weights (Γ)

(4) Completely specified (A) and unscaled.

The covariance matrix of the latent variables (ϕ) has two restrictions.

(1)The latent variables are uncorrelated i.e. ϕ is a diagonal matrix.

(2)The latent variables are correlated i.e. ϕ is a symmetric matrix.

The matrix of errors can take on one of two forms.

(1)The errors are heterogeneous, a general diagonal matrix.

(2)The errors are homogeneous ($\sigma^2 I$).

These models cover Bock and Bargmann's three cases and also a number of Joreskog's confirmatory factor analysis models.

Wiley,Schmidt and Bramble (1975) developed the maximum likelihood estimators for eight of these models. They used only restrictions 3 and 4 for Λ.

2. Multilevel Covariance Structure Models

Bock's model lacks the ability to separate classroom effects from individual effects for data gathered in a natural classroom setting. If tests are given to classes, the variance-covariance matrix of these tests will be affected by the classroom effect.

Assuming classes were sampled at random, Schmidt's Multivariate Random Model gives estimates of the population variance-covariance matrices of the tests at both class and individual level. An individual's score is composed of a number of parts.

(2.1b) $$\underline{y}_{ij} = \underline{\mu} + \underline{\theta}_i + \underline{\alpha}_{ij}$$

where $\underline{y}_{ij}$ is a vector of p scores for person j in group i

$\underline{\mu}$ is a vector of p grand means

$\underline{\theta}_i$ is a vector of p effects due to being in group i

$\underline{\alpha}_{ij}$ is a vector of p effects for person j in group i

The covariance matrix of this model would be :

(2.2b) $$\Sigma_y = Z + \Theta$$

Σ_y is the variance-covariance of p measures for y

Z is the variance-covariance due to class effects

Θ is the variance-covariance due to individual effect.

The two variance-covariance matrices contain information about the class level effects and the individual level effects. These two matrices can be expressed as a function of matrices relating observed to latent

variables and to errors of measurement.

In order to estimate the underlying latent covariance structure and error that compose the two matrices, Schmidt (1971) applied Bock's model to multilevel data. The model included class effects.

(2.3b) $$y_{ijkn} = \mu_k + a_i + b_{ij} + c_{ik} + d_{ijk} + e_{ijkn}$$

where i=1,...,m groups, j=1,...,n(i) students/group, k=1,...,p measures, n=1,...,N students, μ_k is the overall mean of the kth variable, a_i is the effect of group i, b_{ij} is the effect of person j in group i, c_{ik} is an interaction between measure k and class i, d_{ijk} is the interaction between student i in group j and measure k, and e_{ijkn} is measurement error for person j in group i on test k for this occasion.

Notice that the effect at the class level occurs in two terms

(2.4b) $$\theta_{ik} = a_i + c_{ik}$$

and the effect at individual level is found in another two terms.

(2.5b) $$\xi_{jk} = b_{ij} + d_{ijk}$$

Substituting these variables in the model give the following equation.

(2.6b) $$y_{ijkn} = u_k + \theta_{ik} + \xi_{jk} + e_{ijkn}$$

Assuming that u, θ , ξ and e are uncorrelated, the covariance matrix of the y's is given by

(2.7b) $$\Sigma_y = \Omega + \tau + \Psi$$

where Ω is the covariance matrix of $\underline{\theta}$, τ is the covariance matrix of $\underline{\xi}$ and Ψ is the covariance matrix of $\underline{e}$. The model now has three components, Ω which contains the effects at class level, τ which contains the effects at individual level and Ψ which contains the measurement errors.

The interactive random variables $\underline{\theta}_{ki}$ and $\underline{\xi}_{jk}$ could be visualized as combinations of some latent random vectors $\underline{\omega}$ and $\underline{\alpha}$.

(2.8b) $$\underline{\theta}_{ki} = \Lambda_a \underline{\omega}$$

(2.9b) $$\underline{\xi}_{jk} = \Lambda \underline{\alpha}$$

The error matrix Ψ can be rewritten as the linear combination of two components, a within (Ψ_w) and a between group matrix (Ψ_a). From these two assumptions the covariance matrix of $\underline{y}$ is

(2.10b) $$\Sigma_y = \Lambda_a \Phi_a \Lambda_a' + \Lambda \Phi \Lambda' + \Psi_w + \Psi_a$$

where Λ is a pxr matrix of weights relating the observed mean level variables to the vector of r latent variables.

This implies that the basic model for the structural analysis of covariance component matrices of the multivariate random model is given by

(2.11b) $$\Sigma_a = \Lambda_a \Phi_a \Lambda_a + \Psi_a$$

(2.12b) $$\Sigma = \Lambda \Phi \Lambda + \Psi_w$$

where Λ_a is matrix of weights relating the observed mean-level variables to the latent variables $\underline{\omega}$, Λ is matrix of weights relating the observed individual variables to the latent variables $\underline{\alpha}$, Φ_a and Φ are the covariance matrices of the between and within latent effects $\underline{\omega}$, and $\underline{\alpha}$, and Ψ_a, and Ψ_w are the diagonal covariance matrices of the two error matrices. Each of these variance-covariance matrices, Σ_a and Σ, correspond to those considered by Joreskog (1967). The primary difference is that these models, which represent a set of equations, are themselves intended to be simultaneously estimated.

A class of models can be generated by varying the restrictions on the six parameter matrices of this model, Λ_a, Λ, Φ_a, Φ, Ψ_a and Ψ_w. The classifications proposed by Wiley (see last section) can be fit to these parameters.

The two forms of matrices that the latent variance covariance matrices, Φ_a and Φ, can assume are:

(1) The latent variables are uncorrelated i.e. Φ_G is a diagonal matrix.
(2) The latent variables are correlated i.e. Φ_G is a symmetric matrix.

The two error matrices, Ψ_a and Ψ_w, can have one of the following two structures:

(1)The errors are heterogeneous, a general diagonal matrix.
(2)The errors are homogeneous ($\sigma^2 I$).

The four forms of restriction on Λ are

(1) General (Λ) - all elements are to be estimated.

(2) General (Λ) - most elements are to estimated except for certain specified ones

If Λ is reparameterized into Γ A where Γ is a matrix of scaling factors, two different sets of restriction can be applied.

(3) Completely specified (A) and scaled by an unknown but estimable matrix of scaling weights (Γ).

(4) Completely specified (A) and unscaled.

There are (4 x 4 x 2 x 2 x 2 x 2) 256 possible models which can be formed from them.

3. Extension of Φ_a and Φ to causal models.

Once the latent covariance matrices, Φ and Φ_a, are estimated, the matrices themselves can be used to test causal models. Specifically once scaling factors have been specified and measurement error removed, these residual covariance components contain all of the relevant information necessary to analyze a given data structure and hypothesized causal models can be tested.

Joreskog (1971) has developed a procedure for estimating the parameters of causal models using maximum likelihood estimation (LISREL). His procedure estimates the parameters for two components of casual models, namely the measurement model (based on his work mentioned in section A) and the structural model. The measurement model estimates the underlying latent constructs of the model while the structural model specifies the causal relationships among the latent variables. These two components, in turn, are used to describe the causal effects and the amount of unexplained variance among the observed variables.

Wisenbaker (1980) extended Joreskog's model to multilevel situations. His work simultaneously estimated parameters of causal models at both the between and within levels.

The focus in this dissertation is on the estimation of Φ_a and Φ. One natural extension of this work is to develop the algorithm necessary for directly estimating the parameters of Wisenbaker's causal model when groups are unbalanced.

CHAPTER III: REVIEW OF RELEVANT WORK

1. Schmidt's Structural Model

The focus of this dissertation is the estimation of the latent covariance matrices Φ_a and Φ. Schmidt (1969) developed a general procedure for estimating these latent covariance matrices for multivariate normal data. Assuming classes to be drawn at random, the random multivariate model is :

$$\underline{y}_{ij} = \underline{y}_{..} + \underline{a}_i + \underline{e}_{ij} \qquad (3.1a)$$

where $\underline{y}_{ij}$ is the observed set of individual level variables for p values and $\underline{y}_{..}$ is a p x 1 vector of general means. The term $\underline{a}_i$ is a random vector of schools and $\underline{e}_{ij}$ is a random vector of errors. Both of these are considered to be distributed multivariately normal with zero mean vectors and covariance matrices Σ_a and Σ_e. This would imply that the covariance structure for this model would be:

$$\Sigma_y = \Sigma_a + \Sigma_e \qquad (3.2a)$$

Usually Σ_a and Σ_e are estimated by using the expectations of the mean squares of a Multivariate Analysis of Variance (MANOVA). In the random multivariate model it can be shown that

(3.3a) $E(S(w)/[kn-k]) = \Sigma_e$

and

(3.4a) $E(S(b)/[k-1]) = \Sigma_e + n \Sigma_a$.

By using these formulae, Σ_a and Σ_e can be estimated from the following equations

(3.5a) $\hat{\Sigma}_e = S(w)/[kn-k]$

(3.6a) $\hat{\Sigma}_a = (1/n)\{ S(b)/[k-1] - S(w)/[kn-k] \}$.

Unfortunately this method can yield non-positive definite estimates of the matrix Σ_a.

Schmidt used the principle of maximum likelihood to estimate these two matrices. The data in a random model would consist of m factor levels each containing n subjects with p measures on each subject.

		Dependent Variables (measures)				
	Subjects					
Factor		1	2	3	. . .	p
Levels	1					
	2					
1	.					
	.					
	n					

	1
	2
2	.
	.
	n
---	---
	.
.	.
.	.
.	.
	.
---	---
	1
	2
m	.
	.
	n
---	---

Basing the likelihood function on the general notions of Tiao and Tan, (1965), the data can be visualized as m independent observations from a np-dimensional multivariate normal distribution. The general linear model for any y is given by

$$\underline{y} = 1 \otimes \underline{u} + 1 \otimes \underline{a}_i + \underline{e} \qquad (3.7a)$$

where $1 \otimes \underline{u}$ is a vector of pn means (p means repeated over n times),

$\underline{a}_i$ is vector of pn effects (p effects repeated over n times) and $\underline{e}$ is a vector of pn errors. Both $\underline{a}_i$ and $\underline{e}$ are considered to have come from multivariate normal distributions

(3.8a) $$\underline{a} \sim N(0, \Sigma_a) \qquad \underline{e} \sim N(0, \Sigma_e)$$

The covariance matrix for this model is

(3.9a) $$\Sigma_y = \Sigma_{np} = 11' \otimes \Sigma_a + I \otimes \Sigma_e$$

This appears as

$$\begin{bmatrix} \Sigma_a + \Sigma_e & \cdots & \Sigma_a \\ \Sigma_a & \cdots & \\ \Sigma_a & \cdots & \Sigma_a + \Sigma_e \end{bmatrix}$$

The covariance between observations within a factor level is given by

(3.10a) $$\text{Cov}(\underline{Y}_j, Y_i') = \Sigma_a \qquad i \neq j$$

The density function of y is then

(3.11a) $$f(y) = (2\pi)^{-np/2} |\Sigma_{np}|^{-1/2} \exp\{-\tfrac{1}{2}[(\underline{y} - 1\otimes\underline{u})'\Sigma_{np}^{-1}(\underline{y} - 1\otimes\underline{u})]\}$$

from which the likelihood function follows

(3.12a) $$L(\mu, \Sigma_{np}) = (2\pi)^{-mnp/2}|\Sigma_{np}|^{-m/2}\exp\{-\tfrac{1}{2}[\textstyle\sum_{i=1}^{m}(y_i - 1\otimes\underline{u})'\Sigma_{np}^{-1}(y_i - 1\otimes\underline{u})]\}$$

The matrix Σ_{np} must be expressed in terms of Σ_a and Σ_e. The relationship is given in (3.9a), from which the following inverse and determinate follow.

(3.13a) $$|\Sigma_{np}| = |\Sigma_e|^{n-1} + |\Sigma_e + n\Sigma_a|$$

(3.14a) $$\Sigma_{np}^{-1} = I \otimes \Sigma_e^{-1} - 11' \otimes (\Sigma_e + n\Sigma_a)^{-1} \Sigma_a^{-1} \Sigma_e^{-1}$$

The Likelihood can be simplified as

(3.14a) $$L(\mu, \Sigma_a, \Sigma_e) = (2\pi)^{-mnp/2}|\Sigma_e|^{(m-mn)/2}\exp\{-\tfrac{1}{2}[\operatorname{tr}\{\Sigma_e^{-1}S\} + m\operatorname{tr}\{(\Sigma_e + n\Sigma_a)^{-1}S_a\} + mn\operatorname{tr}\{(\Sigma_e + n\Sigma_a)^{-1}(\bar{y} - \mu)(\bar{y} - \mu)'\}]\}$$

where $$S_e = 1/mn \sum_{j=1}^{n}\sum_{i=1}^{m}(y_{ij} - y_i)(y_{ij} - y_i)'$$

$$S_a = n/m \sum_{i=1}^{m}(y_i - y)(y_i - y)'$$

and y_{ij} is a pxl observation vector for the jth person in the ith group. The log of the likelihood is

(3.15a) $$\operatorname{Log}(L(\mu, \Sigma_a, \Sigma_e) = -\tfrac{mnp}{2}\log(2\pi) + \tfrac{n-mn}{2}\log|\Sigma_e| - \tfrac{m}{2}\log|\Sigma_e + n\Sigma_a| - \tfrac{1}{2}[mn\operatorname{tr}\{\Sigma_e^{-1}S\} + m\operatorname{tr}((\Sigma_e + n\Sigma_a)^{-1}S_a) + mn\operatorname{tr}\{(\Sigma_e + n\Sigma_a)^{-1}(y-u)(y-u)'\}]$$

The effective part of the log likelihood function for the estimation of Σ_a and Σ_e is given by

$$(3.16a)\quad \text{Log } L(\mu, \Sigma_a, \Sigma_e) = -\frac{m-mn}{2}\log|\Sigma_e| - \frac{m}{2}\log|\Sigma_e + n\Sigma_a| - \frac{mn}{2}\text{tr}\{\Sigma_a^{-1}S\} - \frac{m}{2}\text{tr}\{(\Sigma_a + n\Sigma_e)^{-1}S_a\}\ .$$

By expressing Σ_{np} in this manner, Schmidt was able to obtain the following maximum likelihood estimates for Σ_e and Σ_a.

$$(3.17a)\qquad \hat{\Sigma}_e = [n/(n-1)]S_e$$

$$(3.18a)\qquad \hat{\Sigma}_a = \frac{1}{m}(S_a - [n/(n-1)]S_e)$$

This gives estimates of the between and with-in covariance matrices but says nothing about the latent constructs or the measurement error associated with the observed values. The equation for a single observation with latent constructs is

$$(3.19a)\qquad y_{ijkn} = u_k + a_i + b_{ij} + c_{ik} + d_{ijk} + e_{ijkn}$$

i=1,...,m groups j=1,...,n(i) students/group k=1,...,p measures
n=1,...,N students

where u is the mean of the kth variable, a is the effect of group i, b is the effect of person j in group i, c is an interaction between measure k and class i, d is the interaction between student i in group j and measure k, and e is measurement error for person j in group i on test k for this occasion. Notice that the class effect occurs in two terms.

(3.20a) $$\underline{\theta}_{ik} = a_i + c_{ik}$$

The effect due to individuals exists in two terms.

(3.21a) $$\underline{\xi}_{jk} = b_{ij} + d_{ijk}$$

The model can be written in terms of the effects at each level.

(3.22a) $$\underline{y}_{ijkn} = \underline{u}_k + \underline{\theta}_{ik} + \underline{\xi}_{jk} + \underline{e}_{ijkn}$$

Assuming that $\underline{u}$, $\underline{\theta}$, $\underline{\xi}$ and $\underline{e}$ are uncorrelated, the covariance matrix of the y's are

(3.23a) $$\Sigma_y = \Omega + \tau + \Psi$$

where Ω is the covariance matrix of $\underline{\theta}$, τ is the covariance matrix of ξ and Ψ is the covariance matrix of $\underline{e}$. The model now has three components, Ω which contains the effects at class level, τ which contains the effects at individual level and Ψ which contains the measurement errors.

The vectors $\underline{\theta}$ and $\underline{\xi}$ could be visualized as combinations of latent random vectors $\underline{\omega}$ and $\underline{\alpha}$.

(3.24a) $$\underline{\theta}_{ki} = \lambda_a \underline{\omega}$$

(3.25a) $$\underline{\xi}_{jk} = \lambda \underline{\alpha}$$

The error matrix Ψ can be rewritten as the linear combination of two components, a within (Ψ_w) and a between group matrix (Ψ_a). From these two assumptions the covariance matrix of $\underline{y}$ is

(3.26a) $$\Sigma_y = \Lambda_a \Phi_a \Lambda_a' + \Lambda \Phi \Lambda' + \Psi_w + \Psi_a$$

where Λ is a pxr matrix of weights relating the observed mean level variables to the vector of r latent variables.

This implies that the basic model for the structural analysis of covariance component matrices of the multivariate random model is

(3.27a) $$\Sigma_a = \Lambda_a \Phi_a \Lambda_a + \Psi_a$$

(3.28a) $$\Sigma = \Lambda \Phi \Lambda + \Psi_w$$

where Λ_a is matrix of weights relating the observed mean-level variables to the latent variables $\underline{\omega}$, Λ is matrix of weights relating the observed individual variables to the latent variables $\underline{\alpha}$, Φ_a and Φ are the covariance matrices of the between and within latent effects $\underline{\omega}$, and $\underline{\alpha}$, and Ψ_a, and Ψ_w are the diagonal covariance matrices of the two error matrices.

Substituting the structural model for Σ_a and Σ_e into the likelihood function in (3.16a) gives the maximum likelihood appropriate for the structural analysis. Taking partial derivitives of the log likelihood in respect to Φ_a, Φ, Λ_a, Λ, Ψ_a, and Ψ_w and setting them equal to zero will yield maximum likelihood estimates of those parameter matrices. These equations proved to be to complicated too be solved

algebraically and Schmidt used the modified method of Davidson as an algorithm to estimate the matrices.

Formulating the maximum likelihood equation by considering the data as $\underline{m}$ random vectors from a multivariate normal distribution of $\underline{np}$ measures constrained the model to have the same number of individuals in each group (i.e. there must be $\underline{p}$ measures for $\underline{n}$ students). When groups have unequal numbers of students, the likelihood function developed in (3.16a) will no longer hold true. In education, researchers are often in the position of collecting data for groups of unequal sizes. To obtain maximum likelihood estimates of the structural matrices in this situation requires either a new analytic strategy or the development of an alternative likelihood function. However finding maximum likelihood estimates of the covariance matrices of an unbalanced design has proved to be difficult.

2. Unbalanced Designs

In multivariate analysis very little has been done to exam the effects of unequal group size on the estimation of the covariance matrix, although there has been much exploration in estimating variance components in the this design for the univariate case. Since Anderson (1984) feels that a number of statistical problems arising in multivariate populations are straightforward analogs of problems arising in univariate populations and the suitable method for handling these problems are similar; parsimony would suggest looking at previous developments regarding the univariate case.

Searles (1971) points out the following problems which must be faced when dealing with unbalanced designs:

> "The property of unbiasedness itself merits questioning in the case of variance component estimators. This is so because with unbalanced data from random models the concept of repetitions of similarly structured data and associated repetitions of estimators is often not appropriate --- more data, maybe, but not necessarily with the same pattern of unbiasedness. Replications of data can not be thought of as mere resamplings of the data already available."

and

> "even in the simplest of cases the effect of the n-pattern on properties of estimators is apparently itself a function

of the variance components being estimated. The effects of unbalancedness therefore appear to differ according to the values of the true variance components."

This last statement refers to the fact that the MINQUE procedure and those based on it rely on the researcher choosing the "true" ratio of the between variance component to the within variance component of the variable.

Welsh (1937) was the first to point out how unequal number of subjects in each group can affect the estimation and testing of statistical hypotheses. Henderson (1953) proposed three methods for estimating variance components for the unbalanced random design, using the expectations of the Random Anova Model.

Rao (1971) advanced a new method for estimating variance components called MINQUE, a minimum quadratic unbiased estimator. Ahrens, Kleffe and Tenzler (1981) state "this procedure provides some kind of optimality and does not refer to the normal assumption" and "MINQUE ... has been justified by heuristic arguments without reference to the normal distribution". Formulas for the MINQUE have been developed with increasing explicitness by Lamotte (1973, 1976) and Ahrens (1978). MINQUE has also been developed for more difficult designs (e.g. see Kleffe (1977)) .

MINQUE can at times give negative estimates of the variance components. Rao (1972), in turn, developed MINQE which gives variance estimates that are always positive but can be biased. It may be noted that no properties are yet known about this estimator.

Searle (1972) devotes an entire review to the methods of variance

estimation in unbalanced random designs. The estimators reviewed are all unbiased, their other properties are unknown. Most of these estimation procedures can lead to negative estimates of the variance component.

Chatterjee and Das (1983) developed a simple estimator of variance components in the random model based on Weighted Least Squares (WLS). They found that as the number of classes increase the proposed estimator is seen as not only to be the best asymtoptically normal but also to be asymtoptically equivalent to the maximum likelihood estimates. A review of recent developments in WLS can be found in Williams, Radcliffe and Speed (1975).

There is no agreement on what constitutes a good estimator of the covariance when groups are unbalanced. As shown above there are many different measures each with its own strengths and weaknesses.

CHAPTER IV: STATEMENT OF PROBLEM

The interest of this dissertation lies in the latent covariance structure implied by the simple true score model. Based on the Simple Multivariate Random Effects Model, the two variance components, between (Σ_a) and within (Σ), are expressed as linear combinations of a set of latent variables.

$$(4.1) \quad \Sigma_a = \Lambda_a \Phi_a \Lambda_a + \Psi_a$$

$$(4.2) \quad \Sigma = \Lambda \Phi \Lambda + \Psi$$

It is the latent covariance matrices Φ_a and Φ that are of primary interest. In chapter 3, a maximum likelihood procedure developed by Schmidt was presented for estimating the error matrices Ψ_a and Ψ and the latent covariance matrices Φ_a and Φ when Λ_a and Λ are known. However, this procedure can only be used when groups in the sample contain the same number of individuals. If the number of individuals in each group is different, this estimation procedure would not be directly appropriate.

The focus of this dissertation is upon the estimation of the group and individual level variances, with measurement error removed, when group sizes vary in a sample.

A promising approach is the EM Algorithm. Developed as an estimation procedure for handling data sets with missing data, it offers a method of finding maximum likelihood estimates of parameters in situations where classical maximum likelihood procedures fail.

The applicability of the EM Algorithm to latent structure models is demonstrated in the next chapter.

CHAPTER V: ESTIMATION PROCEDURE

1. Expectation-Maximization (EM) Algorithm

The EM algorithm has gradually evolved as a method for estimating the parameters of a model when a sample contains missing data. Early works by Hartley (1958), Healy and Westmoratt (1956), Baum et al (1970), and Brown (1974) among others contained specific uses of the EM algorithm under different names. Dempster, Laird and Rubin (1975) developed a more general form for the algorithm and provided a formal proof that if the algorithm did converge, it would result in maximum likelihood estimates.

Missing data cannot be directly measured but exists as function of observed data. This could be censored or truncated data where the value of the data is not the direct value of interest or it could be viewed as being comprised of combinations of latent constructs which form the observed data (Hartly and Hocking, 1971).

Assuming a sample, y, is drawn from a population of a known distribution with unknown parameters ϕ, then y (incomplete data) can be pictured as a subsample of x (complete data) determined by the equation $y = y(x)$. The complete data situation has a family of sampling densities $f(x|\phi)$ depending on ϕ, from which the corresponding family of sampling distributions for the incomplete data, $g(y|\phi)$ can be derived.

The EM algorithm is aimed at finding the ϕ which maximizes $g(y|\phi)$ given an observed y, but making essential use of the family $f(x|\phi)$. There are many possible $f(x|\phi)$ that will generate a $g(y|\phi)$, making the choice of $f(x|\phi)$ a major problem.

Each iteration of the EM algorithm goes through two steps, the expectation step (E-Step) and the maximization step (M-Step). If the complete data, x, comes from a distribution with parameter ϕ, the steps can be stated as follows:

1. E-step: Estimate the complete data sufficient statistics conditional upon the incomplete data, y, and the parameter ϕ. This step provides the connection between the complete data, x, and the incomplete data, y.
2. The M-step determines the parameter ϕ that maximizes the conditional complete data sufficient statistics. This requires writing the Maximum Likelihood equation for ϕ in terms of the complete data.

The sufficient statistics for the complete data are calculated using the incomplete data and estimates of the parameters. (For the first iteration these values of the parameters are given by the user.) The sufficient statistics are then used to estimate the parameters. This value is used to recalculate the sufficient statistics which in turn are used to recalculate the parameter ϕ. The iterations continue until some chosen criterion for convergence is met.

2. Theory for the Restricted Model

It is the application of the E-M algorithm to the estimation of the variance-covariance matrices of a latent multivariate model that is the thrust of this dissertation. The model chosen was based on Schmidt's latent multivariate model with two modifications. First, the groups may or may not contain different numbers of subjects; Schmidt's model allowed only groups of equal size. This modification, however, makes Schmidt's estimation procedure inapplicable. Second, the group level error term in Schmidt's model is not included in the present model.

To estimate the parameters of this unbalanced model, the EM algorithm was employed. The E-step requires the derivation of the conditional sufficient statistics and the M-step requires the Maximum Likelihood Equations of the parameters for the complete data.

The model of interest has the following structure

$$(5.1) \qquad \underline{Y}_{ij} = \underline{\mu} + \lambda_a \underline{\theta}_i + \lambda \, \underline{\alpha}_{ij} + \underline{\epsilon}_{ij}$$

where $\underline{Y}_{ij}$ is a p x 1 vector of observed variables for subject j in group i (incomplete data)

$\underline{\mu}$ is a p x 1 vector of grand means for p variables. (Complete data)

λ_a is a p x q matrix connecting the p observed measures for individuals to the q underlying group latent values.

$\underline{\theta}_i$ is a q x 1 vector of q (where $q \leq p$) latent group effects for group i. (Complete data)

λ is a p x r matrix connecting the p observed measures for individuals to the r underlying individual latent values.

$\underline{\alpha}_{ij}$ is a r x 1 vector of r (where $r \leq p$) latent individual effects for person j in group i. (complete data)

$\underline{\epsilon}_{ij}$ is a p x 1 vector of random error.

For purposes of the derivation of the conditional equations necessary for this EM Algorithm, μ will be considered to be a random vector from a multivariate normal distribution with a mean vector of zero and covariance matrix, Σ_u. Later in the derivation, Σ_u^{-1} will be defined as a zero matrix, yielding posterior estimates of the grand means. This procedure is mentioned in Dempster et al (1976) and further elaborated in Raudenbush (1986).

The latent effects and the error are assumed to come from the multivariate normal distributions

(5.2) $$\underline{\mu} \sim N(\underline{0}, \Sigma_u) \qquad \underline{\alpha} \sim N(\underline{0}, \Phi)$$
$$\underline{\theta} \sim N(\underline{0}, \Phi_a) \qquad \epsilon \sim N(\underline{0}, \Psi) .$$

It is the estimation of Φ_a, Φ and Ψ that is of interest.

Assuming the latent effects are independent, then:

$$
\begin{aligned}
&\mathrm{Cov}(\underline{\mu}, \underline{\theta}) = 0 \qquad \mathrm{Cov}(\underline{\theta}, \underline{\alpha}) = 0 \\
(5.3)\quad &\mathrm{Cov}(\underline{\mu}, \underline{\alpha}) = 0 \qquad \mathrm{Cov}(\underline{\theta}, \underline{\varepsilon}) = 0 \\
&\mathrm{Cov}(\underline{\mu}, \underline{\varepsilon}) = 0 \qquad \mathrm{Cov}(\underline{\alpha}, \underline{\varepsilon}) = 0 .
\end{aligned}
$$

Before finding the conditional sufficient statistics for the maximum likelihood equations, it is important to have a clear understanding of which variables comprise the missing data, the complete data and the parameters of interest. The missing data is our observed dataset $\underline{Y}$. The complete data consists of the three latent variables $\underline{\mu}$, $\underline{\theta}$ and $\underline{\alpha}$. The parameters of interest are the covariance matrices Φ_a, Φ and Ψ.

E-Step

Development of the conditional expectations and dispertions for $\underline{\mu}$, $\underline{\theta}$ and $\underline{\alpha}$ are delineated in this section. These expectations are conditional on the observed data Y and the three parameter matrices Φ_a, Φ and Ψ. The observed dataset has the following expression:

$$
(5.4)\qquad \underline{Y} = 1_N \otimes \underline{\mu} + (X \otimes \lambda_a)\,\underline{\theta} + (I_N \otimes \lambda)\,\underline{\alpha} + \underline{\varepsilon}.
$$

where 1_N is a N x 1 vector and X is an N x m pattern matrix containing 1's and 0's. This matrix connects person N(i) with group m(k).

The variance of the observed dataset, Y, is:

$$\Sigma_y = 11'_N \otimes \Sigma_u + XX' \otimes \lambda_a \Phi_a \lambda_a + I \otimes \lambda \Phi \lambda + I \otimes \Psi \tag{5.5}$$

The joint distribution of $\underline{Y}$, $\underline{\mu}$, $\underline{\theta}$ and $\underline{\alpha}$ is:

$$\begin{bmatrix} Y \\ \mu \\ \alpha \\ \theta \end{bmatrix} \sim N \left[\begin{array}{cccc} \Sigma_y & & \text{Symmetric Matrix} & \\ 1'_N \otimes \Sigma_u & N\Sigma_u & & \\ X' \otimes \Phi_a \lambda'_a & 0 & I_k \otimes \Phi_a & \\ I \otimes \Phi\lambda & 0 & 0 & I_N \otimes \Phi \end{array} \right] \tag{5.6}$$

with

$$\begin{aligned} \text{Cov}(Y, \underline{\mu}) &= 1_N \otimes \Sigma_\mu \\ \text{Cov}(Y, \underline{\theta}) &= X \otimes \Phi_a \\ \text{Cov}(Y, \underline{\alpha}) &= I \otimes \Phi \\ \text{Cov}(\underline{\mu}, \underline{\theta}) &= 0 \\ \text{Cov}(\underline{\mu}, \underline{\alpha}) &= 0 \\ \text{Cov}(\underline{\theta}, \underline{\alpha}) &= 0. \end{aligned}$$

By defining the matrices and vectors as

$$Z = \begin{bmatrix} 1_N \otimes I & X \otimes \lambda_a & I \otimes \lambda \end{bmatrix}$$

$$T = \begin{bmatrix} \mu \\ \theta \\ \alpha \end{bmatrix} \qquad \Omega = \begin{bmatrix} N\Sigma_u & \text{Symmetric Matrix} & \\ 0 & I_k \otimes \Phi_a & \\ 0 & 0 & I_N \otimes \Phi \end{bmatrix} \tag{5.7}$$

the joint normal prior distribution can be written as

$$(5.8)\qquad \begin{pmatrix} \underline{Y} \\ \underline{T} \end{pmatrix} \sim N \begin{pmatrix} Z\,\Omega\,Z' + I \otimes \Psi & Z\,\Omega \\ \Omega\,Z' & \Omega \end{pmatrix}$$

Raudenbush (1986) derived the formulas for the conditional expectation and dispersion of matrices written in this form. The conditional expectation and dispersion of T given Y are:

$$(5.9)\qquad E(\underline{T}|\underline{Y}) = (\,Z'(I \otimes \Psi)^{-1}Z + \Omega^{-1})^{-1}Z'(I \otimes \Psi)^{-1}\underline{Y}$$

$$(5.10)\qquad D(\underline{T}|\underline{Y}) = (\,Z'(I \otimes \Psi)^{-1}Z + \Omega^{-1})^{-1}$$

By substituting the original values of Z, $\underline{T}$ and Ω in (6.7) and allowing Σ_{μ}^{-1} to become a matrix of zeros as previously mentioned, the dispersion matrix becomes

$$(5.11)\qquad D(\underline{T}|\underline{Y}) = \begin{bmatrix} N\,\Psi^{-1} & & \text{Symmetric Matrix} \\ 1_k \otimes n_i\lambda'_a\Psi^{-1} & I_k \otimes (n_i\,\lambda'_a\Psi^{-1}\lambda_a + \Phi_a^{-1}) & \\ 1_N \otimes \lambda'\Psi^{-1} & X \otimes \lambda'_a\Psi^{-1}\lambda & I_N \otimes (\lambda'\Psi^{-1}\lambda + \Phi^{-1}) \end{bmatrix}$$

Partitioning this matrix into a 2 x 2 matrix and applying the procedure in Morrison (1972) for inverting such a matrix, leads to the following values for the elements

$$D(T|Y) = \begin{bmatrix} 1 & & \\ 2 & 3 & \\ 4 & 5 & 6 \end{bmatrix}$$

$$(5.12)\qquad 1 = D(\underline{\mu}|\underline{Y}) = W$$

$$(5.13)\qquad 2 = D(\underline{\Theta}, \mu|\underline{Y}) = -1_k \otimes \Phi_a\lambda_a Q_i W$$

$$(5.14)\quad 3 - D(\underline{\theta}|\underline{Y}) = I_k \otimes (\Phi_a - \Phi_a\lambda_a'Q_i\lambda_a\Phi_a) + 1_k1_k' \otimes \Phi_a\lambda_a'Q_iW\,Q_j\lambda_a\Phi_a$$

$$(5.15)\quad 4 - D(\underline{\alpha}, \mu|\underline{Y}) = -1_N \otimes \Phi\,\lambda'Q_iW$$

$$(5.16)\quad 5 - D(\underline{\alpha}, \theta|\underline{Y}) = 1_N1_K' \otimes (1/n_i)\,\Phi\,\lambda'Q_jW\,Q_i\lambda_a\Phi_a - X \otimes (1/n_i)\,\Phi\,\lambda'Q_jW\,Q_i\lambda_a\Phi_a$$

$$(5.17)\quad 6 - D(\underline{\alpha}|\underline{Y}) = 1_N1_N' \otimes (1/n_i)(1/n_j)\,\Phi\,\lambda\,Q_jW\,Q_i\lambda\,\Phi + I_N \otimes (\Phi - \Phi\,\lambda'M\,\lambda\,\Phi) + XX' \otimes (1/n_i)\Phi\lambda'(M - (1/n_i)Q_i)\lambda\Phi$$

where $M = (\lambda\,\Phi\,\lambda' + \Psi)^{-1}$

$Q_i = (\lambda_a\Phi_a\lambda_a' + (1/n_i)M^{-1})^{-1}$

$W = [\Sigma_{i=1}^{k} Q_i]^{-1}$

The conditional expectations of μ, $\underline{\theta}$ and $\underline{\alpha}$ can be formulated by substituting (6.7) into (6.10). This yields:

$$(5.18)\quad E\begin{bmatrix} \mu|Y \\ \underline{\theta}|Y \\ \underline{\alpha}|Y \end{bmatrix} = \begin{bmatrix} \mu^* \\ \underline{\theta}^* \\ \underline{\alpha}^* \end{bmatrix} - \begin{matrix} \Sigma_{i=1}^{k} W\,Q_i\bar{Y} \\ 1_k \otimes \Phi_a\lambda_a'Q_i(\bar{\underline{Y}}_i - \mu^*) \\ 1_N \otimes \Phi\,\lambda'M(\underline{Y}_{ij} - \lambda\,\theta^* - \mu^*) \end{matrix}$$

M-Step.

The second step of the EM Algorithm requires expressing the maximum likelihood equations of the parameters Φ_a, Φ and Ψ in terms of the complete data. Assuming the values μ, $\underline{\theta}$ and $\underline{\alpha}$ are known, the expression of the likelihoods for our three parameters can be directly stated as:

$$(5.19)\quad L(Y, \Phi_a) = \Pi_{i=1}^{k} (2\pi)^{-p/2}|\Phi_a|^{-1/2} \exp\{-\tfrac{1}{2}\,\theta'\Phi_a^{-1}\theta\}$$

$$(5.20)\quad L(Y, \Phi) = \prod_{i=1}^{N} (2\pi)^{-p/2}|\Phi|^{-1/2} \exp\{-\tfrac{1}{2}\alpha'\Phi^{-1}\alpha\}$$

$$(5.21)\quad L(Y, \Psi) = \prod_{i=1}^{N} (2\pi)^{-p/2}|\Psi|^{-1/2} \exp\{-\tfrac{1}{2}\epsilon'\Psi^{-1}\epsilon\}$$

The maximum likelihood estimate of Φ_a is derived by first finding the log of (5.18).

$$(5.22)\quad \log(L(Y, \Phi_a)) = -kp/2\,(\log(2\pi)) - \tfrac{k}{2}\log|\Phi_a| - \tfrac{1}{2}\textstyle\sum_{i=1}^{K}\theta'\Phi_a^{-1}\theta$$

Taking the derivitive with respect to Φ_a and setting it equal to zero yields:

$$(5.23)\qquad \hat{\Phi}_a = k^{-1}\textstyle\sum_{i=1}^{k}\theta\,\theta'$$

In reality, only the conditional expectation of μ, $\underline{\theta}$ and $\underline{\alpha}$ are known (U^*, θ^*, α^*). By rewriting (5.22) as

$$\hat{\Phi}_a = \tfrac{1}{k}\textstyle\sum_{i=1}^{k}(\underline{\theta}^* + \theta - \theta^*)(\theta^* + \theta - \theta^*)'$$

$$\hat{\Phi}_a = \tfrac{1}{k}\textstyle\sum_{i=1}^{k}(\underline{\theta}^*\underline{\theta}^{*\prime}) + (\theta^*)(\theta - \theta^*)' + (\theta - \theta^*)(\theta^*)' + (\theta - \theta^*)(\theta - \theta^*)'$$

$$(5.24)\qquad \hat{\Phi}_a = k^{-1}\textstyle\sum_{i=1}^{k}(\theta^*\underline{\theta}^{*\prime}) + D(\underline{\theta}^*)\quad .$$

The equation can be clarified by substituting By substituting (5.18) for θ^* and (5.12) for $D(\theta^*)$. The maximum likelihood estimate becomes:

$$(5.25)\quad \hat{\Phi}_a = \Phi_a - \tfrac{1}{k}\textstyle\sum_{i=1}^{k}[\Phi_a\lambda_a'(Q_i - Q_i(A - W)Q_i)\lambda_a\Phi_a]$$

where $A = (\bar{Y}_i - \mu^*)(\bar{Y}_i - \mu^*)'$

Following the same likelihood procedure from steps (5.22) through (5.24) above, gives the following maximum likelihood estimate for Φ:

$$(5.26) \quad \hat{\Phi} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \alpha_{ij}^{*}{}'\alpha_{ij}^{*} + D(\alpha_{ij})$$

Equation (5.25) can be clarified by substituting By substituting (5.18) for α^{*} and (5.15) for $D(\alpha^{*})$. The maximum likelihood estimate becomes:

$$(5.27) \quad \hat{\Phi} = \Phi - \Phi\lambda'[(1/N)\sum_{i=1}^{N}\Phi\lambda'((1/n_i)[Q_i - Q_i(A-W)Q_i] - M[B-(n_i - 1)M^{-1}]M)\lambda\Phi$$

where $B = (\underline{Y}_{ij} - \lambda_a\underline{\theta}^{*} - \mu^{*})(\underline{Y} - \lambda_a\underline{\theta}^{*} - \underline{\mu}^{*})'$

The maximum likelihood for Ψ followed steps (5.22) and (5.23). The maximum likelihood estimate is:

$$(5.28) \quad \hat{\Phi} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \epsilon_{ij}\epsilon'_{ij}$$

Replacing ϵ by $Y - (\mu + \lambda + \lambda\alpha)$ permits (5.25) to be rewritten as:

$$(5.29) \quad \hat{\Psi} = \frac{1}{N}\sum\sum (Y_{ij} - (\mu + \lambda_a\theta + \lambda\alpha))(Y_{ij} - (\mu + \lambda_a\theta + \lambda\alpha))'$$

replacing conditional values expanding the equation yields

$$(5.30) \quad \hat{\Psi} = \frac{1}{N}\sum\sum (Y_{ij} - (\mu^{*}+\lambda_a\theta^{*}+\lambda\alpha^{*}))(Y - (\mu^{*}+\lambda_a\theta^{*}+\lambda\alpha^{*}))' + D(\mu^{*} + \lambda_a\theta^{*} + \lambda\alpha^{*})$$

$$(5.31) \quad D(\mu^{*}+\lambda_a\theta^{*}+\lambda\alpha^{*}) = D(\mu^{*})+2D(\mu^{*}, \lambda_a\theta^{*})+2D(\mu^{*}, \lambda\alpha^{*})+D(\lambda_a\theta^{*}) + 2D(\lambda_a\theta^{*}, \lambda\alpha^{*})+D(\lambda\alpha^{*})$$

By substituting (5.12), (5.13), (5.14), (5.15), (5.16), (5.17) and (5.18) into (5.30) the estimate of Ψ can be written as:

$$(5.32)\quad \Psi = \Psi - \frac{1}{N}\sum_{i=1}^{K}\Psi((1/n_i)[Q_i - Q_i(A-W)Q_i] - M[B-(n_i - 1)M^{-1}]M)\Psi$$

where $B = (\underline{Y}_{ij} - \lambda\,\Theta^{*} - \mu^{*})(\underline{Y} - \lambda_a\underline{\Theta}^{*} - \underline{\mu}^{*})'$

3. The E-M Algorithm.

The implementation of this algorithm is now complete for this latent model. The E-step uses the observed dataset Y and starting values for the parameters Φ_a, Φ and Ψ to find estimates of the following sufficient statistics in (5.12) through (5.18). The M-step finds estimates of the three parameters using the values from the first step in (5.25), (5.27) and (5.32). These estimates are used in step 1 to reestimate the sufficient statistics in (5.12) through (5.18). Then in the M-step, Φ_a, Φ and Ψ are estimated again using the new values from E-step. The algorithm iterates between these two steps until some criteria is reached.

Chapter VI: Design of Study

1. Design.

By applying the estimation procedure (described in Chapter Five) to a set of data sampled from a population of known parameters, a check was provided for the solution of the EM algorithm together with the identification of its properties. Although the underlying parameters of the sample data were known, this was not intended to be a simulation study but, rather, an example of the algorithm's ability to estimate the parameters of a latent model.

The EM algorithm, in operational terms, was used to estimate covariance components from both unbalanced and balanced samples drawn from the same multivariate normal distribution with known parameters. The balanced case contained 30 subjects for each group, while the unbalanced case averaged 30 subjects per group. The distribution of subjects across the groups in the unbalanced case was as follows: 20% of the groups included 10 subjects within each group, 20% had 20 subjects, 20% had 30 subjects, 20% had 40 subjects and, finally, the last 20% had 50 subjects.

The estimates of the balanced and unbalanced samples were both studied while varying two factors, namely the number of groups in the sample (the size) and the particular model being estimated (i.e. the unrestricted model, the correctly restricted model and incorrectly restricted model). The size of the sample consisted of two levels. The small sample consisted of 25 groups and the large sample of 100 groups. The difference in the number of classes gave an indication of the

properties of the EM Algorithm when the sample size varied from 25 groups with 750 subjects to 100 groups with 3000 subjects.

Each of the three models studied contained different sets of parameters to be estimated. The first set involved the two covariance components for the simple multivariate random model (Unrestricted model); Σ_a, the between groups covariance and Σ, the within groups covariance matrix. The Unrestricted model gave estimates of the between and within covariance matrices of the multivariate random model.

The second set consisted of Φ_a, the latent group covariance matrix, Φ, the latent individual covariance matrix and Ψ, the error covariance matrix from the latent multivariate model. These parameters were derived from a latent measurement model based on the multivariate random model. The latent group covariance matrix, Φ_a, and the latent individual covariance matrix, Φ, were allowed to be full rank (i.e. covariances were not constrained to zero) while the error covariance matrix was constrained to a diagonal matrix.

The last set of parameters were from an incorrectly specified latent multivariate random model. The parameters included were Φ^*_a, the latent group covariance matrix, Φ^*, the latent individual covariance matrix and Ψ^*, the error covariance matrix. All three matrices were restricted to diagonal matrices. Applied to data from a population in which the parameter matrices contained non-zero covariance terms, this model demonstrated the reaction of the EM algorithm to incorrectly specified models.

Figure 1 is a diagram of the design. There were 12 cells, each containing 10 replications.

FIGURE 1

Design of Study

Model	Number of Classes: 25 Classes, Balanced	25 Classes, Unbalanced	100 Classes, Balanced	100 Classes, Unbalanced
Unrestricted	a	b	c	d
Correctly Specified	e	f	g	h
Incorrectly Specified	i	j	k	l

1. Each cell contains 10 repetitions (different sets of data)
2. Cells a through f contain comparable datasets; the same can be said for datasets g through l.

2. Generation of Data.

Implementation of the experimental design required a method of creating samples drawn from a population of known parameters. The data had to fit the assumptions for the multivariate random latent model specified in Chapter Five. The values of the parameters chosen for this example are listed in Table 1.

Each subject's four observed scores, $\underline{y}_{ij}$, were a combination of three latent group effects, $\underline{\theta}_{i.}$, three latent individual effects, $\underline{\omega}_{ij}$, four measurement errors, $\underline{\epsilon}_{ij}$, and four grand means, $\underline{\mu}_{..}$. The most direct way to generate a dataset of observed values containing these characteristics is to create four separate vectors, one for each effect, for each subject and then to create the observed values through the equation

6.1 $$\underline{y}_{ij} = \underline{\mu}_{..} + \Lambda_a \underline{\theta}_{i.} + \Lambda \underline{\omega}_{ij} + \underline{\epsilon}_{ij}$$

This is the equation from the random latent model in Chapter Five. Unlike the other three vectors, the grand mean, $\underline{\mu}$, will be identical for all subjects. Each vector is representative of a sample vector from a normal distribution with mean zero and variance covariance matrix as shown in Table 1.

The SAS package contains a subroutine which generates independent values from a univariate normal distribution with mean of zero and variance of one. By repeated applications three vectors of dimensions 3 x 1, 3 x 1 and 4 x 1 were created for each subject. The vectors, X($\underline{\theta}$), X($\underline{\omega}$) and X($\underline{\epsilon}$) each constitute a random sample of values

Table 1

Parameter Values Used in Data Generation

The dimension of the observed variables is four (p=4). The dimension of the latent group variables is three (r=3) and the dimension of the latent individual variables is three (s=3).

The pattern matrices connecting the latent to the observed variables are:

$$LA = \lambda_a = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 1 & 0.5 & -0.5 \\ 1 & -0.5 & 0.5 \\ 1 & -0.5 & -0.5 \end{bmatrix} \qquad L = \lambda = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 1 & 0.5 & -0.5 \\ 1 & -0.5 & 0.5 \\ 1 & -0.5 & -0.5 \end{bmatrix}$$

The latent error, individual and group matrices are:

$$TH = \Phi_a = \begin{bmatrix} 64 & 8 & 40 \\ 8 & 5 & 7 \\ 40 & 7 & 107 \end{bmatrix} \qquad OM = \Phi = \begin{bmatrix} 25 & 10 & 15 \\ 10 & 20 & 10 \\ 15 & 10 & 35 \end{bmatrix}$$

$$PS = \Psi = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 12 \end{bmatrix}$$

The Expected values of the grand means of the four observed variables were set to zero ($\underline{\mu} = \underline{0}$).

drawn from multivariate normal distributions having means of 0 and identity covariance matrices.

The three vectors X($\underline{\theta}$), X($\underline{\omega}$), and X($\underline{\varepsilon}$) are each created from multivariate normal random distributions with mean vectors of zero and identity variance-covariance matrices. Sample data from a population with that parameter were obtained by multiplying each vector by the cholesky of the known parameter matrix. The final sample vectors were

$$X'(\underline{\theta}) = T(\Phi_a) * X(\underline{\theta})$$

$$X'(\underline{\omega}) = T(\Phi) * X(\underline{\omega})$$

$$X'(\underline{\varepsilon}) = T(\Psi) * X(\underline{\varepsilon})$$

where $T(\Phi_a)$ is the cholesky of Φ_a, $T(\Phi)$ is the cholesky of Φ and $T(\Psi)$ is the cholesky of Ψ.

By using $X'(\underline{\theta})$, $X'(\underline{\omega})$, $X'(\underline{\varepsilon})$ and $\underline{\mu}$ together as shown in equation 6.1, an observed sample data set from a population of known latent parameters was created. Each data set was created through the SAS normal random generation procedure using different seed numbers. The data sets were then used by a computer program to estimate the values of the parameters.

3. Starting Values of Parameter Matrices

The EM algorithm requires starting values for each parameter being estimated. The computer program written for this study estimates two sets of parameters, first the unrestricted between and within covariance matrices for the observed data, then in turn, the parameters of a more restricted latent model. The calculations of the starting values for the parameters of the latent model are based on the final estimates of the unrestricted between and within covariance matrices.

Starting values for the between and within variance covariance matrices were estimated from equations 6.2 and 6.3 as developed by Schmidt (1971).

$$6.2 \qquad \hat{\Sigma} = [n/(n-1)]S$$

$$6.3 \qquad \hat{\Sigma}_a = 1/m\ (S_a - [n/(n-1)]S)$$

These estimates are maximum likelihood estimators under the case of equal group size.

These equations are used as starting values for an unbalanced design with one modification, replacing n by its harmonic mean. In a balanced design the use of harmonic $\underline{n}$ will give the maximum likelihood estimates. When the design is unbalanced, the harmonic $\underline{n}$ will give weighted starting values for the parameters.

After Σ and Σ_a have been estimated, starting values for the parameters of the latent model Φ_a, Φ and Ψ are found. Assuming $\Sigma_a = \Lambda_a \Phi_a \Lambda_a$ and $\Sigma = \Lambda\ \Phi\ \Lambda + \Psi$ then

$$6.4 \qquad \hat{\Phi}_a = (\Lambda'\Lambda_a)^{-1}\Lambda'\Sigma\Lambda_a(\Lambda'\Lambda_a)^{-1}$$

$$6.5 \qquad \hat{\Phi} = (\Lambda'\Lambda)^{-1}\Lambda\,\Sigma\,\Lambda\,(\Lambda'\Lambda)^{-1}$$

$$6.6 \qquad \hat{\Psi} = \Sigma_a - \Lambda_a\hat{\Phi}\,\Lambda_a + \Sigma - \Lambda\,\hat{\Phi}\,\Lambda$$

These values form the starting estimates of the parameters for the latent model.

4. The Computer Program.

The computer program was written using the SAS procedure "Proc Matrix".

1. Necessary Input: Y - a N x p matrix for p measures on N individuals, Λ_a and Λ_a a which are pattern matrices connecting the underlying latent variable with the observed-level variables and a K x 2 matrix, Z, which specifies the number of students in each group.
2. Next the program creates two new matrices, YM, a K x p matrix of means for each group, and SS, a Kp x p matrix containing the sum of squares for each group.

Estimate of the parameters of a simple random model.

3. Using Schmidt's Maximum Likelihood Equations 6.2 and 6.3 the program then estimates starting values for Σ_a and Σ.

E-step

4. Using the equations given above the program next estimates the conditional varibles W, Q, $\underline{\mu}^*$, and $\underline{\theta}^*$.

M-step

5. Using these values the program then recalculates estimates for the parameters Σ_a and Σ.

6. A reiteration then occurs between steps 4 and 5 until the changes in Σ_a and Σ are less than 0.01

Estimate the parameters of the latent Restricted Model.

7. The next step in the program computes the starting estimates of Φ_a, Φ and Ψ from Σ_a and Σ from the two parameter matrices in step 5 and estimates the sufficient statistics W, Q, M, $\underline{\theta}^*$ and $\underline{\mu}^*$ from these values (E-step).

M-step

8. Reestimated values for the parameters Φ_a, Φ and Ψ are then obtained and, finally, the program iterates between steps 7 and 8 until the changes in the parameters Φ_a, Φ and Ψ are less than 0.01.

CHAPTER VII: RESULTS

1. Design and Measures.

The EM Algorithm's ability to estimate covariance components in both balanced and unbalanced latent multivariate random effects models were demonstrated by estimating the parameters of independent samples generated from a population with a known distribution. The balanced samples contained 30 subjects for each group, and the unbalanced samples averaged 30 subjects per group.

The estimates of the balanced and unbalanced samples were studied across two dimensions, namely the number of groups in the sample and the type of model being estimated (i.e. the unrestricted model, the correctly specified latent model and an incorrectly specified latent model. Twenty elements were estimated in the unrestricted model, 10 for the Phi matrix and 10 for the Psi matrix. Sixteen elements were estimated in the correctly specified model; six for the Ph matrix, six for the Om matrix and four for the Ps matrix. The incorrectly specified model differed from the correctly specified model only in the number of matrix elements being estimated. Only the 10 diagonal elements were estimated, three for the Ph matrix, three for the Om matrix and four for the Ps matrix.

Tables 2, 3 and 4 contain descriptive statistics of the estimates of the individual items of the covariance matrices for the three models over 10 repetitions for different situations. These tables include the Expected Value of the parameter (E), the Mean, the Standard Deviation (SD), the Bias, the Mean Square Error (MSE), the Bias divided by the

parameter's Expected Value (B/E), Ratio(1) and Ratio(2). Mean and Standard Deviation are self explanatory. The Bias is the difference between the Expected value of the parameter and its sample mean. The Mean Square Error (MSE) is the averaged squared deviation of the parameter estimates around their Expected value.

The ratio of the Bias to the Expected Value (B/E) of a variance or covariance term converts the Bias into a percentage of the parameter's Expected Value, giving it a relative value.

The difference between the MSEs of the balanced sample estimate and the corresponding unbalanced sample estimate divided by the MSE of the unbalanced sample estimate (Ratio(1)) yields a comparison of the MSEs of the two types of datasets.

The last measure (Ratio(2)) is the difference between the MSEs of an element of the incorrectly specified model and the correctly specified model divided by the MSE of the element of the correctly specified model. The ratio gives a comparison of the precision of the two models.

The three tables contain the lower diagonal elements of the different covariance matrices. In Tables 2 and 3, the elements of the latent matrix at group level, Φ_a, are labeled Ph(X), the elements of the latent individual level matrix, Φ , are Om(X) and the elements of the error matrix from the latent models, Ψ ,are labeled Ps(X). The (X) corresponds to the elements position in the lower diagonal. The latent covariance matrices would be

$$Ph(\ ) = \begin{bmatrix} 1 & & \\ 2 & 3 & \\ 4 & 5 & 6 \end{bmatrix} \qquad Om(\ ) = \begin{bmatrix} 1 & & \\ 2 & 3 & \\ 4 & 5 & 6 \end{bmatrix} \qquad Ps(\ ) = \begin{bmatrix} 1 & & & \\ & 2 & & \\ & & 3 & \\ & & & 4 \end{bmatrix}$$

Ph(3) is the variance of the second latent measure and Ph(5) is the covariance between the second and third elements of the Ph matrix. Ps(2) would the variance of the second observed measure.

The elements of the unrestricted model are similarly labeled. The elements of the between group covariance matrix, Σ_a, are labeled Phi(X) while the elements of the within covariance matrix, Σ , are Psi(X). They are the same dimension as the error matrix of the latent model, 4 x 4, but include the six covariance terms in their estimates.

Tables 5 through 7 contain aggregated statistics for each matrix in the latent models under the different conditions. Table 8 lists the Maximum Likelihood Ratio test of the estimates of the correctly and incorrectly specified latent models. Tables 9 and 10 list aggregated statistics for each matrix in the unrestricted models under the different conditions. Table 11 has information about the iterations necessary for the algorithm to converge.

With only 10 repetitions per cell, the power of any statistical test would be low. Although some characteristics of the estimation procedure may appear with this size sample, it should be recalled that this was just a demonstration of the use of the EM algorithm for an unbalanced latent model under different circumstances and not a statistical study.

2. Results of Estimation Procedure

Table 2 contains information about the estimates of the parameters for four different situations when the datasets are comprised of 100 groups. The four conditions are: (1) applying the correctly specified model to the balanced samples; (2) applying the correctly specified model to the unbalanced samples; (3) applying the incorrectly specified model to the balanced samples; and, lastly, (4) applying the incorrectly specified model to the unbalanced samples. The items in this table represent the statistics for cells g, h, k and l in Figure 1.

The B/E in Table 2 indicates the percentage of bias of the estimates. The correctly specified latent model had values of the B/E ranging from -0.9% to -13.7% for the estimates of the elements of the Ph matrix for the balanced data and 15.1% to -11.5% for the unbalanced dataset. The B/E of the estimates of the elements of the Om matrix ranged from 8.7% to -1.3% for the balanced data and from 10.9% through -9.8% for the unbalanced data and the Ps matrix had B/E values ranging from 8.7% to -27.4% for the balanced data and 5.2% to -41.5% for the unbalanced data.

In the incorrectly specified model, the estimates of the elements of the Ph matrix, the variance components, had B/E values are almost identical to the corresponding elements in the correctly specified model. The B/E of the estimates of the elements of the Om matrix ranged from 6.4% to -10.4% for the balanced data and 6.6% through -10.2% for the unbalanced data and the Ps matrix had B/E values

TABLE 2

Summary Statistics of the Latent Models for Balanced and Unbalanced Data Sets with 100 Groups

Ph Matrix

		Expected	Balance (h)	Unbalance (k)	Balance Diagonal (i)	Unbalance Diagonal (l)
Ph(1)	Mean	64.000	63.400	62.960	63.440	63.240
	SD		9.330	10.710	9.290	10.530
	MSE		87.449	115.906	86.653	111.523
	Bias(B)		-0.600	-1.040	-0.560	-0.760
	B/E		-0.009	-0.016	-0.009	-0.012
	Ratio 1			0.325		0.287
	Ratio 2				-0.009	-0.038
Ph(2)	Mean	8.000	7.600	9.210	0.000	0.000
	SD		1.880	2.090		
	MSE		3.712	5.995		
	Bias(B)		-0.400	1.210		
	B/E		-0.050	0.151		
	Ratio 1			0.615		
	Ratio 2					
Ph(3)	Mean	5.000	4.550	4.450	4.550	4.450
	SD		0.640	0.600	0.650	0.560
	MSE		0.635	0.696	0.648	0.650
	Bias(B)		-0.450	-0.550	-0.450	-0.550
	B/E		-0.090	-0.110	-0.090	-0.110
	Ratio 1			0.097		0.003
	Ratio 2				0.020	-0.067
Ph(4)	Mean	40.000	39.060	38.720	0.000	0.000
	SD		6.270	5.610		
	MSE		40.295	33.293		
	Bias(B)		-0.940	-1.280		
	B/E		-0.023	-0.032		
	Ratio 1			-0.174		
	Ratio 2					
Ph(5)	Mean	7.000	6.040	6.560	0.000	0.000
	SD		3.210	3.360		
	MSE		11.328	11.505		
	Bias(B)		-0.960	-0.440		
	B/E		-0.137	-0.063		
	Ratio 1			0.016		
	Ratio 2					

TABLE 2 (Continued)

Ph Matrix

		Expected	Balance	Unbalance	Balance Diagonal	Unbalance Diagonal
Ph(6)	Mean	107.000	98.420	99.560	98.190	99.210
	SD		14.980	12.760	15.470	13.060
	MSE		306.196	224.322	325.561	237.990
	Bias(B)		-8.580	-7.440	-8.810	-7.790
	B/E		-0.080	-0.070	-0.082	-0.073
	Ratio 1			-0.267		-0.269
	Ratio 2				0.063	0.061

Om Matrix

Om(1)	Mean	25.000	25.580	25.900	26.600	26.640
	SD		0.880	0.990	1.000	0.960
	MSE		1.148	1.880	3.844	3.910
	Bias(B)		0.580	0.900	1.600	1.640
	B/E		0.023	0.036	0.064	0.066
	Ratio 1			0.637		0.017
	Ratio 2				2.348	1.080
Om(2)	Mean	10.000	10.870	11.090	0.000	0.000
	SD		0.830	0.900		
	MSE		1.530	2.130		
	Bias(B)		0.870	1.090		
	B/E		0.087	0.109		
	Ratio 1			0.392		
	Ratio 2					
Om(3)	Mean	20.000	19.740	19.990	18.780	18.750
	SD		0.830	0.830	0.710	0.650
	MSE		0.764	0.689	2.158	2.159
	Bias(B)		-0.260	-0.010	-1.220	-1.250
	B/E		-0.013	-0.001	-0.061	-0.063
	Ratio 1			-0.098		0.000
	Ratio 2				1.824	2.133
Om(4)	Mean	15.000	15.020	14.840	0.000	0.000
	SD		1.300	1.410		
	MSE		1.690	2.017		
	Bias(B)		0.020	-0.160		
	B/E		0.001	-0.011		
	Ratio 1			0.193		
	Ratio 2					

TABLE 2 (Continued)

Om Matrix

		Expected	Balance	Unbalance	Balance Diagonal	Unbalance Diagonal
Om(5)	Mean	10.000	10.390	9.020	0.000	0.000
	SD		3.710	4.140		
	MSE		13.933	18.207		
	Bias(B)		0.390	-0.980		
	B/E		0.039	-0.098		
	Ratio 1			0.307		
	Ratio 2					
Om(6)	Mean	35.000	36.300	36.370	31.370	31.440
	SD		2.900	3.180	0.670	0.690
	MSE		10.288	12.198	15.090	14.558
	Bias(B)		1.300	1.370	-3.630	-3.560
	B/E		0.037	0.039	-0.104	-0.102
	Ratio 1			0.186		-0.035
	Ratio 2				0.467	0.193

Ps Matrix

Ps(1)	Mean	5.000	3.630	5.260	12.120	12.070
	SD		4.410	5.460	0.460	0.470
	MSE		21.534	29.887	56.539	55.760
	Bias(B)		-1.370	0.260	7.120	7.070
	B/E		-0.274	0.052	1.424	1.414
	Ratio 1			0.388		-0.014
	Ratio 2				1.626	0.866
Ps(2)	Mean	6.000	5.250	3.510	2.070	2.050
	SD		3.650	4.170	0.260	0.250
	MSE		13.948	24.278	17.229	17.399
	Bias(B)		-0.750	-2.490	-3.930	-3.950
	B/E		-0.125	-0.415	-0.655	-0.658
	Ratio 1			0.741		0.010
	Ratio 2				0.235	-0.283
Ps(3)	Mean	11.000	12.280	11.530	6.260	6.250
	SD		3.060	3.540	0.570	0.550
	MSE		11.184	12.844	25.289	25.372
	Bias(B)		1.280	0.530	-4.740	-4.750
	B/E		0.116	0.048	-0.431	-0.432
	Ratio 1			0.148		0.003
	Ratio 2				1.261	0.975

TABLE 2 (Continued)

Om Matrix

		Expected	Balance	Unbalance	Balance Diagonal	Unbalance Diagonal
Ps(4)	Mean	12.000	13.040	13.890	15.390	15.480
	SD		2.120	1.960	0.610	0.610
	MSE		5.696	7.811	13.141	13.828
	Bias(B)		1.040	1.890	3.390	3.480
	B/E		0.087	0.158	0.283	0.290
	Ratio 1			0.371		0.052
	Ratio 2				1.307	0.770

ranging from 142% to -66% for the balanced data and 141% to -66% for the unbalanced data.

Table 3 contains information about the estimates of the parameters for the four different situations in Table 2 when the datasets are comprised of 25 groups. Statistics for cells e, f, i and j in Figure 1 are presented in this table.

The B/E of the corresponding items in Table 3 show higher percentages of bias than those in Table 2. The correctly specified latent model had values of B/E ranging from -6.6% to -33.9% for the estimates of the elements of the Ph matrix for the balanced data and -0.6% to -33.9% for the unbalanced dataset. The B/E of the estimates of the elements of the Om matrix ranged from 7.9% to 1.1% for the balanced data and 5.7% through -1.7% for the unbalanced data and the Ph matrix had B/E values ranging from 14.3% to -43.6% for the balanced data and 3.7% to -13.0% for the unbalanced data.

In the incorrectly specified model the estimates of the elements of the Ph matrix, the variance components, had B/E values almost identical to the corresponding elements in the correctly specified model. The B/E of the estimates of the elements of the Om matrix ranged from 5.1% to -8.0% for the balanced data and 4.9% to -8.4% for the unbalanced data and the Ps matrix had B/E values ranging from 136% to -65% for the balanced data and 136% to -64% for the unbalanced data.

Table 4 contains information about the estimates of the variables of the unrestricted model under the following conditions: (1) for balanced samples containing 100 groups; (2) for unbalanced samples containing 100 groups; (3) for balanced samples containing 25 groups; and lastly (4) for unbalanced samples containing 25 groups. Although

TABLE 3

Summary Statistics of the Latent Models for Balanced and Unbalanced Data Sets with 25 Groups

Ph Matrix

		Expected	Balance (e)	Unbalance (f)	Balance Diagonal (i)	Unbalance Diagonal (j)
Ph(1)	Mean	64.000	59.750	63.620	59.890	64.070
	SD		16.730	20.280	16.790	16.830
	MSE		299.962	411.439	300.673	283.254
	Bias		-4.250	-0.380	-4.110	0.070
	B/E		-0.066	-0.006	-0.064	0.001
	Ratio 1			0.372		-0.058
	Ratio 2				0.002	-0.312
Ph(2)	Mean	8.000	6.260	7.290	0.000	0.000
	SD		3.050	3.520		
	MSE		12.667	12.951		
	Bias		-1.740	-0.710		
	B/E		-0.218	-0.089		
	Ratio 1			0.022		
	Ratio 2					
Ph(3)	Mean	5.000	4.040	4.400	4.060	4.130
	SD		1.580	1.530	1.580	2.080
	MSE		3.520	2.741	3.478	5.167
	Bias		-0.960	-0.600	-0.940	-0.870
	B/E		-0.192	-0.120	-0.188	-0.174
	Ratio 1			-0.221		0.486
	Ratio 2				-0.012	0.885
Ph(4)	Mean	40.000	28.940	33.070	0.000	0.000
	SD		14.760	18.120		
	MSE		353.773	381.695		
	Bias		-11.060	-6.930		
	B/E		-0.277	-0.173		
	Ratio 1			0.079		
	Ratio 2					
Ph(5)	Mean	7.000	4.660	4.630	0.000	0.000
	SD		4.160	6.170		
	MSE		23.390	44.310		
	Bias		-2.340	-2.370		
	B/E		-0.334	-0.339		
	Ratio 1			0.894		
	Ratio 2					

TABLE 3 (CONTINUED)

Ph Matrix

		Expected	Balance	Unbalance	Balance Diagonal	Unbalance Diagonal
Ph(6)	Mean	107.000	77.450	78.720	77.910	78.770
	SD		22.380	26.990	23.190	27.490
	MSE		1471.089	1617.081	1478.030	1641.181
	Bias		-29.550	-28.280	-29.090	-28.230
	B/E		-0.276	-0.264	-0.272	-0.264
	Ratio 1			0.099		0.110
	Ratio 2				0.005	0.015

Om Matrix

Om(1)	Mean	25.000	25.380	25.580	26.270	26.220
	SD		1.850	1.680	2.130	2.160
	MSE		3.583	3.196	6.329	6.319
	Bias		0.380	0.580	1.270	1.220
	B/E		0.015	0.023	0.051	0.049
	Ratio 1			-0.108		-0.002
	Ratio 2				0.766	0.977
Om(2)	Mean	10.000	10.770	10.540	0.000	0.000
	SD		1.790	1.540		
	MSE		3.863	2.696		
	Bias		0.770	0.540		
	B/E		0.077	0.054		
	Ratio 1			-0.302		
	Ratio 2					
Om(3)	Mean	20.000	21.010	21.150	20.060	20.430
	SD		2.400	2.310	2.230	2.400
	MSE		6.893	6.806	4.977	5.965
	Bias		1.010	1.150	0.060	0.430
	B/E		0.051	0.057	0.003	0.021
	Ratio 1			-0.013		0.199
	Ratio 2				-0.278	-0.123
Om(4)	Mean	15.000	15.170	14.970	0.000	0.000
	SD		1.230	1.610		
	MSE		1.545	2.593		
	Bias		0.170	-0.030		
	B/E		0.011	-0.002		
	Ratio 1			0.678		
	Ratio 2					

TABLE 3 (CONTINUED)

Om Matrix

		Expected	Balance	Unbalance	Balance Diagonal	Unbalance Diagonal
Om(5)	Mean	10.000	10.500	9.830	0.000	0.000
	SD		2.830	3.900		
	MSE		8.287	15.242		
	Bias		0.500	-0.170		
	B/E		0.050	-0.017		
	B /MSE		0.034	0.002		
	Ratio 1			0.839		
	Ratio 2					
Om(6)	Mean	35.000	37.780	35.910	32.190	32.070
	SD		5.340	2.010	2.420	2.370
	MSE		37.103	4.960	14.630	15.156
	Bias		2.780	0.910	-2.810	-2.930
	B/E		0.079	0.026	-0.080	-0.084
	Ratio 1			-0.866		0.036
	Ratio 2				-0.606	2.055

Ps Matrix

Ps(1)	Mean	5.000	2.820	4.350	11.810	11.820
	SD		4.250	4.700	1.450	1.430
	MSE		23.343	22.559	53.632	53.725
	Bias		-2.180	-0.650	6.810	6.820
	B/E		-0.436	-0.130	1.362	1.364
	Ratio 1			-0.034		0.002
	Ratio 2				1.298	1.382
Ps(2)	Mean	6.000	5.260	5.380	2.100	2.190
	SD		3.560	3.560	1.190	1.230
	MSE		13.282	13.101	18.316	17.642
	Bias		-0.740	-0.620	-3.900	-3.810
	B/E		-0.123	-0.103	-0.650	-0.635
	Ratio 1			-0.014		-0.037
	Ratio 2				0.379	0.347
Ps(3)	Mean	11.000	12.570	11.410	6.430	6.520
	SD		1.960	2.810	1.020	0.980
	MSE		6.580	8.083	24.246	23.261
	Bias		1.570	0.410	-4.570	-4.480
	B/E		0.143	0.037	-0.415	-0.407
	Ratio 1			0.228		-0.041
	Ratio 2				2.685	1.878

TABLE 3 (CONTINUED)

Ps Matrix

		Expected	Balance	Unbalance	Balance Diagonal	Unbalance Diagonal
Ps(4)	Mean	12.000	12.280	11.830	14.060	13.880
	SD		2.330	1.290	1.220	0.430
	MSE		5.516	1.696	6.204	4.112
	Bias		0.280	-0.170	2.060	1.880
	B/E		0.023	-0.014	0.172	0.157
	Ratio 1			-0.692		-0.337
	Ratio 2				0.125	1.424

TABLE 4

Summary Statistics of the Unrestricted Models for Balanced and Unbalanced Data Sets for Both 25 and 100 Groups

Phi Matrix

		Expected	100 Groups Balance (c)	100 Groups Unbalance (d)	25 Groups Balance (a)	25 Groups Unbalance (b)
Phi(1)	Mean	143.500	141.962	141.230	127.691	139.317
	SD		16.359	16.865	34.203	44.422
	MSE		270.245	290.154	1447.539	1992.756
	Bias(B)		-1.538	-2.270	-15.809	-4.183
	Bias/Expected(B/E)		-0.011	-0.016	-0.110	-0.029
	Ratio 1			0.074		0.377
	Ratio 2				4.356	5.868
Phi(2)	Mean	46.500	48.521	47.472	51.845	56.904
	SD		12.099	13.831	20.467	23.513
	MSE		150.924	192.346	450.641	673.131
	Bias		2.021	0.972	5.345	10.404
	B/E		0.043	0.021	0.115	0.224
	Ratio 1			0.274		0.494
	Ratio 2				1.986	2.500
Phi(3)	Mean	32.500	35.633	34.445	39.816	44.156
	SD		11.939	12.684	19.156	21.152
	MSE		153.446	165.087	426.423	598.365
	Bias		3.133	1.945	7.316	11.656
	B/E		0.096	0.060	0.225	0.359
	Ratio 1			0.076		0.403
	Ratio 2				1.779	2.625
Phi(4)	Mean	49.500	49.980	49.060	53.453	53.344
	SD		7.474	7.330	8.875	9.751
	MSE		56.117	53.944	96.128	111.500
	Bias		0.480	-0.440	3.953	3.844
	B/E		0.010	-0.009	0.080	0.078
	Ratio 1			-0.039		0.160
	Ratio 2				0.713	1.067
Phi(5)	Mean	129.500	128.490	128.139	116.448	126.424
	SD		14.652	15.129	33.778	42.182
	MSE		215.815	230.945	1330.236	1789.834
	Bias		-1.010	-1.361	-13.052	-3.076
	B/E		-0.008	-0.011	-0.101	-0.024
	Ratio 1			0.070		0.346
	Ratio 2				5.164	6.750

TABLE 4 (Continued)

Phi Matrix

		Expected	100 Groups Balance	100 Groups Unbalance	25 Groups Balance	25 Groups Unbalance
Phi(6)	Mean	56.500	55.834	55.339	59.803	60.973
	SD		9.087	8.879	9.528	10.791
	MSE		83.066	80.334	102.905	138.676
	Bias		-0.666	-1.161	3.303	4.473
	B/E		-0.012	-0.021	0.058	0.079
	Ratio 1			-0.033		0.348
	Ratio 2				0.239	0.726
Phi(7)	Mean	120.500	119.725	119.638	109.674	118.120
	SD		13.156	13.843	33.908	40.686
	MSE		173.748	192.454	1279.977	1661.644
	Bias		-0.775	-0.862	-10.826	-2.380
	B/E		-0.006	-0.007	-0.090	-0.020
	Ratio 1			0.108		0.298
	Ratio 2				6.367	7.634
Phi(8)	Mean	39.500	41.359	41.109	45.143	49.719
	SD		10.461	12.173	21.005	23.745
	MSE		113.272	151.058	476.592	679.856
	Bias		1.859	1.609	5.643	10.219
	B/E		0.047	0.041	0.143	0.259
	Ratio 1			0.334		0.426
	Ratio 2				3.207	3.501
Phi(9)	Mean	30.500	33.114	32.664	37.596	41.590
	SD		10.240	10.916	19.657	21.313
	MSE		112.450	124.362	442.346	590.897
	Bias		2.614	2.164	7.096	11.090
	B/E		0.086	0.071	0.233	0.364
	Ratio 1			0.106		0.336
	Ratio 2				2.934	3.751
Phi(10)	Mean	47.500	46.748	47.373	51.063	50.635
	SD		6.523	6.079	8.626	9.364
	MSE		43.178	36.972	88.513	98.605
	Bias		-0.752	-0.127	3.563	3.135
	B/E		-0.016	-0.003	0.075	0.066
	Ratio 1			-0.144		0.114
	Ratio 2				1.050	1.667

TABLE 4 (Continued)

Psi Matrix

		Expected	100 Groups Balance	100 Groups Unbalance	25 Groups Balance	25 Groups Unbalance
Psi(1)	Mean	73.750	73.985	74.058	73.822	73.404
	SD		1.682	1.659	4.438	4.283
	MSE		2.890	2.858	19.702	18.477
	Bias		0.235	0.308	0.072	-0.346
	B/E		0.003	0.004	0.001	-0.005
	Ratio 1			-0.011		-0.062
	Ratio 2				5.816	5.466
Psi(2)	Mean	31.250	31.613	31.616	31.392	31.609
	SD		1.118	1.095	3.445	3.387
	MSE		1.396	1.348	11.890	11.615
	Bias		0.363	0.366	0.142	0.359
	B/E		0.012	0.012	0.005	0.011
	Ratio 1			-0.035		-0.023
	Ratio 2				7.515	7.617
Psi(3)	Mean	6.250	6.533	6.513	6.341	6.817
	SD		0.609	0.576	2.235	2.399
	MSE		0.460	0.409	5.004	6.112
	Bias		0.283	0.263	0.091	0.567
	B/E		0.045	0.042	0.015	0.091
	Ratio 1			-0.111		0.221
	Ratio 2				9.882	13.958
Psi(4)	Mean	13.750	13.925	13.894	14.022	14.362
	SD		0.767	0.728	1.601	1.545
	MSE		0.622	0.553	2.645	2.803
	Bias		0.175	0.144	0.272	0.612
	B/E		0.013	0.010	0.020	0.045
	Ratio 1			-0.111		0.060
	Ratio 2				3.251	4.069
Psi(5)	Mean	43.750	44.095	44.216	43.565	43.741
	SD		1.396	1.323	2.512	2.201
	MSE		2.081	1.992	6.348	4.844
	Bias		0.345	0.466	-0.185	-0.009
	B/E		0.008	0.011	-0.004	0.000
	Ratio 1			-0.043		-0.237
	Ratio 2				2.050	1.432

TABLE 4 (Continued)

Psi Matrix

		Expected	100 Groups Balance	100 Groups Unbalance	25 Groups Balance	25 Groups Unbalance
Psi(6)	Mean	34.750	34.987	34.936	35.350	35.831
	SD		0.857	0.864	2.306	1.922
	MSE		0.797	0.785	5.718	4.992
	Bias		0.237	0.186	0.600	1.081
	B/E		0.007	0.005	0.017	0.031
	Ratio 1			-0.015		-0.127
	Ratio 2				6.175	5.360
Psi(7)	Mean	49.750	49.935	50.034	49.971	50.328
	SD		1.306	1.237	2.009	2.072
	MSE		1.744	1.620	4.090	4.664
	Bias		0.185	0.284	0.221	0.578
	B/E		0.004	0.006	0.004	0.012
	Ratio 1			-0.071		0.140
	Ratio 2				1.346	1.880
Psi(8)	Mean	16.250	16.690	16.710	15.915	16.095
	SD		1.122	1.084	2.278	2.303
	MSE		1.474	1.410	5.314	5.331
	Bias		0.440	0.460	-0.335	-0.155
	B/E		0.027	0.028	-0.021	-0.010
	Ratio 1			-0.043		0.003
	Ratio 2				2.605	2.780
Psi(9)	Mean	11.250	11.354	11.308	11.681	12.121
	SD		0.660	0.628	1.812	1.836
	MSE		0.448	0.398	3.490	4.214
	Bias		0.104	0.058	0.431	0.871
	B/E		0.009	0.005	0.038	0.077
	Ratio 1			-0.111		0.207
	Ratio 2				6.796	9.584
Psi(10)	Mean	30.750	30.860	30.861	30.662	30.770
	SD		0.905	0.878	1.210	1.141
	MSE		0.832	0.785	1.473	1.302
	Bias		0.110	0.111	-0.088	0.020
	B/E		0.004	0.004	-0.003	0.001
	Ratio 1			-0.058		-0.116
	Ratio 2				0.769	0.660

the Unrestricted model was a linear combination of the latent variance and covariance components of the correctly specified model, it was directly estimated from the data.

For the estimates from data containing 100 groups, B/E ranged from 9.6% to -1.6% for the balanced data and 7.1% to -12.7% for the unbalanced data. The estimates from the datasets containing 25 groups had B/E ranging from 23.3% to -11.0% for the balanced data and 36.4% to -2.9% for the unbalanced data.

Table 5 contains the average B/E value for each matrix, for the balanced and unbalanced data, for both the correctly and incorrectly specified latent models.

In the correctly specified latent model, the matrices from large group data (n=100 groups) showed lower average B/E than matrices from small group data (n=25 groups). The Om and Ps matrices had average B/E percentages ranging between 2.9% and -4.9% versus 4.7% and -9.8%. The Ph matrices had values of -2.3% to -6.5% versus -16.5% to -22.7%.

The findings regarding the incorrectly specified model were not as consistent. The B/E for the Om matrix was lower for the small group data in both the balanced and unbalanced datasets. The Ph and Ps matrices, on the other hand, had lower B/E for the large group data.

The data from the balanced group would be expected to have the best estimates with the smallest MSE. If this procedure can get estimates of the unbalanced design, with only a small increase in the MSE, than the estimation procedure would be practical. Table 6 contains the average values of Ratio(1), the ratio of the difference between the MSE of the balanced and unbalanced samples to the MSE of the balanced sample, of each matrix of the two latent models for both sample sizes.

Table 5

Average B/E of the Latent Models for Balanced and Unbalanced Data Sets for Both 25 and 100 Groups

25 Classes

	(e) Balance	(f) Unbalance	(i) Incorrect Balance	(j) Incorrect Unbalance
Ph	-0.227	-0.165	-0.175	-0.146
	(0.093)	(0.121)	(0.130)	(0.130)
Om	0.047	0.024	-0.009	-0.004
	(0.029)	(0.030)	(0.047)	(0.050)
Ps	-0.098	-0.053	0.115	0.120
	(0.250)	(0.078)	(0.899)	(0.894)

100 Classes

	(g) Balance	(h) Unbalance	(k) Incorrect Balance	(l) Incorrect Unbalance
Ph	-0.065	-0.023	-0.060	-0.065
	(0.047)	(0.091)	(0.045)	(0.050)
Om	0.029	0.012	-0.034	-0.033
	(0.035)	(0.069)	(0.087)	(0.088)
Ps	-0.049	-0.026	0.103	0.102
	(0.185)	(0.256)	(0.936)	(0.933)

(a), ... indicates row of figures under letter refer to cell in Figure 1.

Table 6

Average Ratio(1) of the Ph, Om and Ps Matrices for the Latent Models in Data Sets for both 25 and 100 Groups

25 Classes

	Unbalance	Incorrect Unbalance
Ph	0.208 (0.386)	0.179 (0.227)
Om	0.038 (0.634)	0.078 (0.089)
Ps	-0.128 (0.395)	-0.103 (0.113)

100 Classes

	Unbalance	Incorrect Unbalance
Ph	0.102 (0.326)	0.007 (0.197)
Om	0.269 (0.245)	-0.006 (0.019)
Ps	0.412 (0.245)	0.013 (0.028)

Ratio 1 = (MSE(Unbalanced)-MSE(Balanced))/(MSE(Balanced))

In the correctly specified model for the samples containing 100 groups, the unbalanced samples had greater MSE than the balanced in all three of the matrices averaging from 10% to 41% more. For samples containing 25 groups, the two individual level matrices, Om and Ps, showed little or negative increase on the average. The Ps matrix had an average drop of 13% for the MSE, while the Om matrix had only a slight increase of 3%. The Ph matrix had an average increase of 20%.

The incorrectly specified model showed the same results for samples containing 25 groups. The balanced samples containing 100 groups showed very little difference in the MSE from the unbalanced samples containing 100 groups for all three matrices.

The estimates of the balanced samples improved more than the unbalanced samples (in terms of MSE) as group size increases.

The imposition of a structure on the data, in turn, permits specification of incorrect models. Table 7 contains the average values of Ratio(2), the ratio of the difference between the MSE of the correctly and incorrectly specified latent models divided by the MSE of the correctly specified model, of each matrix for the balanced and unbalanced samples for both sample sizes.

In the samples containing 25 groups, the balanced data shows little difference between the MSE's of the incorrectly and correctly specified models for the Ph and Om matrices, 0% and -4%. The Ps matrix had an increase in the MSE of 112%. The unbalance sample had rises in the MSE of 19% for the Ph matrix, 97% for the Om matrix and 125% for the Ps matrix.

For the balanced and unbalanced samples containing 100 groups, the Ph matrix showed little increase in the MSE between the correctly

Table 7

Average Ratio(2) of the Ph, Om and Ps Matrices for Balanced and Unbalanced Data Sets for both 25 and 100 Groups

25 Classes

	Incorrect Balance	Incorrect Unbalance
Ph	-0.002 (0.006)	0.196 (0.619)
Om	-0.039 (0.507)	0.970 (1.093)
Ps	1.121 (1.157)	1.258 (0.647)

100 Classes

	Incorrect Balance	Incorrect Unbalance
Ph	0.025 (0.036)	-0.015 (0.067)
Om	1.546 (0.971)	1.135 (0.971)
Ps	1.107 (0.604)	0.582 (0.583)

Ratio 2 = (MSE(Incorrect)-MSE(Correct))/(MSE(Correct))

and incorrectly specified models, while the Om matrix increased 13% and 15%, respectively, and the Ps matrix showed increases of 58% and 110%.

The large rises in the MSE of the Ps matrices for the incorrectly specified model can be explained by reviewing Tables 2 and 3. The Ps matrix of the incorrectly specified model is very biased with a small sampling variance. It is this bias that causes the MSE to greatly increase. Without knowing the true values of the variances and covariances, the incorrect model would be tempting to accept because of the small sampling variance that accompanies it.

The incorrectly specified model does well in estimating the variances of Ph, but Om and Ps show problems with their estimates. In Tables 1 and 2 the standard deviation of the Ps elements in the correctly specified model vary between 2 to 10 times as large as the corresponding elements for the incorrectly specified model. On the other hand, the bias of the estimates of the elements of the Ps matrix, in the incorrectly specified model, range between 2 to 10 times as large as the bias for the corresponding elements in the correctly specified model. The percentage of the MSE which was due to bias in the incorrectly specified model in Ps (all sizes) ranged between 91% to 99.6%. The low standard deviation of the sample, but very incorrect estimates, indicate a very consistent but extremely biased estimate. The direction of the bias was not consistent across elements.

It is also important to test the model for fit. By using the Maximum Likelihood Ratio (MLR), the correctly and incorrectly specified models can be tested for fit. If the MLR is significant, it is an indication that the model does not fit the data. Table 8 contains the statistics of the maximum likelihood ratio for the correctly and

Table 8

Maximum Likelihood Ratio Test of the Fit of the Correct and Incorrect Models

25 Classes

	* (b) Balance	* (e) Unbalance	** (c) Incorrect Balance	** (f) Incorrect Unbalance
Mean	2.77 (1.76)	4.21 (3.77)	129.77 (39.16)	144.53 (54.53)
Minimum	1.29	1.58	81.12	68.36
Maximum	7.41	13.13	210.99	231.57
No. of samples significant at $p < .05$	(1)	(2)	(10)	(10)

100 Classes

	* (h) Balance	* (k) Unbalance	** (i) Incorrect Balance	** (l) Incorrect Unbalance
Mean	7.41 (7.05)	15.16 (12.49)	618.59 (99.09)	603.88 (89.86)
Minimum	0.82	2.46	475.21	443.02
Maximum	22.72	34.25	769.72	783.49
No. of samples significant at $p < .05$	(3)	(6)	(10)	(10)

*** df = 4**
**** df = 10**

incorrectly specified models. The MLR's of the correctly specified model are lower than those for the incorrectly specified model's by a factor of more than 10. The tests of fit for all forty samples for the incorrectly specified model had significant MLR's. Only 12 of those samples were significant for the correctly specified model, nine of which were from samples containing 100 groups.

The datasets containing 100 groups had MLR's four times the size of those from datasets with 25 groups. If the dataset is very large, the fit may be acceptable, but the MLR significant. This is a common problem in covariate structural analysis. The same problem occurs in Lisrel when using a very large sample.

The unrestricted model estimated only one covariance matrix for the group level and one matrix for the individual level. Neither of these matrices, Phi or Psi, were structured or constrained. This particular model was estimated separately from the latent models.

The starting values of the unrestricted model were Maximum Likelihood Estimates when the data was balanced. Schmidt's Maximum Likelihood Equations for the between and within covariance matrices of a multivariate random model were used as starting points. This algorithm always converged at the end of the first iteration for balanced datasets.

Table 9 contains the average B/E of Phi and Psi in the unrestricted model. Both the balanced and unbalanced samples showed little difference in the average B/E of either matrix when the data contained 100 groups. The B/E averaged less than 2.4% of the expected value. When the data was comprised of 25 groups, the average B/E of the Psi matrix was less than 2.5% for both the balanced and unbalanced

Table 9

Average B/E of the Unrestricted Model for Balanced and Unbalanced Data Sets for Both 25 and 100 Groups

	(c) 100 Balance	(d) 100 Unbalance	(a) 25 Balance	(b) 25 Unbalance
Phi	0.023 (0.042)	0.013 (0.033)	0.063 (0.127)	0.135 (0.154)
Psi	0.013 (0.013)	0.013 (0.013)	0.007 (0.016)	0.025 (0.035)

samples. The Phi matrix, on the other hand, had values of 13% and 17% for the balanced and unbalanced datasets irrespectively. The increase in the sample size, from 750 to 3000, had little effect on the B/E for the Psi matrix. The increase from 25 to 100 groups, however, reduced the average B/E for the estimates of the elements of the Phi matrix in the balanced and unbalanced samples from 13% and 17% to 2% and 1%.

Table 10 summarizes the values of Ratio(1) for the unrestricted model. The Phi matrix showed a higher average MSE for the unbalanced samples than for the balanced samples in both small and large group data. (33% and 8%, respectively). The Psi matrix showed little difference between the MSE's of the balanced and unbalanced samples for samples of either size. As the number of groups in a sample increase, the MSE of the unbalanced data evidently approaches that of the balanced data.

One last note, statistical theory states that as the sample size increases the sample variance will decrease. The SD should be about

Table 10

Average Ratio(1) of the Phi and Psi Matrices for Balanced and Unbalanced Data Sets for Both 25 and 100 Groups

	100 Unbalance	25 Unbalance
Phi	0.083	0.330
	(0.142)	(0.116)
Psi	-0.061	0.007
	(0.039)	(0.151)

twice as large for the 25 group as for the 100 group samples. This is borne out for the Ph and Om matrices, but not by the Ps matrix. The Ps matrix has approximately the same SD for both samples of 25 and 100 groups.

3. Results of the Process.

A section of the findings of this study apply to the process of the EM algorithm Problems encountered in the procedure of the algorithm in this environment may apply to other situations.

Table 11 contains information on the number of iterations the estimation procedure took to converge for each of the cells in Figure 1. The incorrectly specified model needed the most iterations to converge for both data containing 100 groups and data containing 25 groups. The means of the balanced and unbalanced samples were very close, 82 and 86 for the data with 25 groups and 99 and 100 in the data with 100 groups.

The correctly specified model averaged 74 and 76 iterations for the balanced and unbalanced samples of 25 groups. For data containing 100 groups, the unbalanced sample averaged 38 iterations less than the balanced sample, 99 vs 61.

The unrestricted model averaged fewer iterations than either of the latent models for all conditions. The latent models, at the least, averaged over 50 more iterations than the unrestricted model. The samples containing 100 groups for the unbalanced sample averaged less iterations than the unbalanced sample from the 25 group case, 5.8 against 10.3.

The unrestricted model for the balanced sample under both cases always stopped after the first iteration. The starting value for the algorithm was Schmidt's maximum likelihood estimators for the between and within models. This confirmed that the algorithm was capable of stopping at a maximum likelihood estimate.

TABLE 11

Iterations Required by Algorithm to Convergence

	Balanced Design			Unbalanced Design		
	Mean	Standard Deviation	Range	Mean	Standard Deviation	Range
25 Groups						
Unrestricted	1.0	-	1 - 1	10.3	7.79	1 - 28
Correctly Specified	74.6	53.43	23 - 198	76.1	39.40	37 - 159
Incorrectly Specified	82.2	19.42	50 - 105	86.6	16.83	53 - 102
100 Groups						
Unrestricted	1.0	-	1 - 1	5.8	0.42	5 - 6
Correctly Specified	98.8	62.84	27 - 192	61.3	24.87	27 - 117
Incorrectly Specified	99.3	3.62	92 - 104	99.9	2.56	96 - 104

Ph and Om were estimated as full matrices for the correctly specified model and as diagonal matrices for the incorrectly specified model.

Convergence was found to be a problem for two datasets and they were not used in the final analysis. In one dataset under the unbalanced case, the unrestricted model moved toward convergence for seven iterations until only one element of the three covariance matrices was slightly larger than criterion. At iteration eight, the estimates of the parameters of the matrix
diverged from the expected values until the program automatically stopped at the 250th iteration. The estimates of the parameters had significantly diverged from the maximum likelihood values. A slight change in the values of the starting matrix of this dataset caused the algorithm to converge in seven iterations.

In the Unrestricted model the Psi matrix converged very quickly. The convergence of the model came only after the elements in the Phi matrix reached the criteria.

In the correctly specified latent model the Ph matrix was the first matrix for all of the elements to reach the convergence criteria. The Ps matrix was the last in which all elements reached criteria.

Finally, the starting values of the parameters affected the final estimated values at which the algorithm converged. Specifically, proximity of the starting values to the true values appeared to be positively related to how closely the final estimated parameters would be to the maximum likelihood estimates when reaching criterion. Criterion was reached when all elements in the covariance matrices changed less than .01. Using a set of data, an initial computer run was done on the data using values close to the expected values as starting values. These starting values caused the final estimates of the parameters to be close to the Expected value. A second run of the

program was then done on this data using Schmidt's formula to find starting values for the parameters. The run resulted with estimates of the parameter that were not as close to the expected values as were those from the first run. The criteria were ignored and the second was allowed to continue for 95 more iterations. It then reached values which were nearly identical to those from the first run.

CHAPTER VIII: DISCUSSION

1. Summary and Conclusions.

Although statistical procedures are available for estimating treatment effects for students taught in classrooms, these procedures are applicable, only, when every class has the same number of students. The present study investigated a procedure that was originally established to handle missing data (EM Algorithm) but which also provides a solution to the problem of estimating parameters in multivariate analysis when samples contain unequal group sizes. The focus of the present dissertation was on the estimation of latent group and individual level variances and covariances with measurement error removed when group sizes varied in a sample. Previous methods could only find maximum likelihood estimates for this problem if the dataset contained groups of equal size. The EM Algorithm offers a method for finding maximum likelihood estimates of parameters in situations where classical maximum likelihood procedures failed.

To estimate a set parameters, the EM Algorithm requires two steps, an expectation step (E-step) and a maximization step (M-step). The E-step is characterized by the formulation of the sufficient statistics in terms of the observed data and the parameters. The M-step consists of developing the maximum likelihood equations for the parameters in terms of the conditional statistics. Using given starting values for the parameters, the algorithm calculates the sufficient statistics in the E-step. These values are used to estimate the parameters in the M-step. The algorithm returns to the E-step to

recalculate the sufficient statistics based on the new values of the parameters. The parameters are reestimated using these new values of the sufficient statistics. The algorithm iterates between the E-step and the M-step until a specified criteria is reached.

The estimate of balanced and unbalanced samples were both studied while varying two factors, mainly the number of groups in the sample (the size) and the particular model being estimated (that is to say, the unrestricted model, the correctly specified model and the incorrectly specified model). The unrestricted model was

$$\underline{Y}_{ij} = \underline{\mu} + \underline{\gamma}_i + \underline{\epsilon}_{ij} \tag{8.1}$$

$\underline{Y}$ is a p x 1 vector of observed data.

$\underline{\mu}$ is a p x 1 vector of grand means.

$\underline{\gamma}$ is a p x 1 vector of group effects.

$\underline{\epsilon}$ is a p x 1 vector of individual effects.

These variables were considered to have come from multivariate normal distributions:

$$\underline{Y} \sim N(\underline{0}, \Sigma_Y) \qquad \underline{\gamma} \sim N(\underline{0}, \Sigma_\gamma) \tag{8.2}$$
$$\underline{\mu} \sim N(\underline{0}, \Sigma_\mu) \qquad \underline{\epsilon} \sim N(\underline{0}, \Sigma_\epsilon) .$$

The parameters of interest for this model were Σ_γ and Σ_ϵ.

The correctly specified model was visualized as the application of a structure to the unrestricted model. By assuming $\underline{\gamma} = \lambda_a \underline{\theta}$ and $\underline{\epsilon} = \lambda_\alpha + \underline{\epsilon 1}$, (8.1) becomes:

(8.3) $$\underline{Y}_{ij} = \underline{\mu} + \lambda_a \underline{\theta}_i + \lambda \underline{\alpha}_{ij} + \underline{\epsilon 1}_{ij}$$

$\underline{Y}$ is a p x 1 vector of observed data.

$\underline{\mu}$ is a p x 1 vector of grand means.

λ_a is a p x q matrix connecting the p observed variables with q latent group variables.

$\underline{\theta}$ is a q x 1 vector ($q \leq p$) of the latent group effects.

λ is a p x r matrix connecting the p observed variables with r latent group variables.

$\underline{\alpha}$ is a r x 1 vector ($r \leq p$) of the latent individual effects.

$\underline{\epsilon 1}$ is a p x 1 vector of the latent individual errors.

with $\underline{Y} \sim N(\underline{0}, \Sigma_Y)$ $\quad \underline{\theta} \sim N(\underline{0}, \Phi_a)$

$\underline{\mu} \sim N(\underline{0}, \Sigma_\mu)$ $\quad \underline{\alpha} \sim N(\underline{0}, \Phi)$

$\underline{\epsilon 1} \sim N(\underline{0}, \Psi)$.

The parameters of interest in this model are Φ_a, Φ and Ψ.

The incorrectly specified model differs from the correctly specified model in the constraints placed on the two latent covariance matrices. In the incorrect model, the latent covariance matrices, and Φ_a and Φ are considered to be diagonal matrices. No such constraints are placed on the latent matrices in the correct model.

Tests of the model based on the criteria of convergence showed this estimation procedure to be a satisfactory and effective method in theory. However, once the study had been completed, it was recognized that a model containing a group error term would be a necessity for all

practical applications. This model can be described as follows:

$$\underline{Y}_{ij} = \underline{\mu} + \lambda_a \underline{\theta}_i + \lambda \underline{\alpha}_{ij} + \underline{\epsilon 1}_{ij} + \underline{\epsilon 2}_{ij} \tag{8.5}$$

$\underline{Y}$ is a p x 1 vector of observed data.

$\underline{\mu}$ is a p x 1 vector of grand means.

λ_a is a p x q matrix connecting the p observed variables with q latent group variables.

$\underline{\theta}$ is a q x 1 vector ($q \leq p$) of the latent group effects.

λ is a p x r matrix connecting the p observed variables with r latent group variables.

$\underline{\alpha}$ is a r x 1 vector ($r \leq p$) of the latent individual effects.

$\underline{\epsilon 1}$ is a p x 1 vector of the latent individual errors.

$\underline{\epsilon 2}$ is a p x 1 vector of the latent group errors.

with

$$\underline{Y} \sim N(\underline{0}, \Sigma_Y) \qquad \underline{\gamma} \sim N(\underline{0}, \Phi_a)$$

$$\underline{\mu} \sim N(\underline{0}, \Sigma_\mu) \qquad \underline{\alpha} \sim N(\underline{0}, \Phi)$$

$$\underline{\epsilon 1} \sim N(\underline{0}, \Psi_1) \qquad \underline{\epsilon 2} \sim N(\underline{0}, \Psi_2).$$

The parameters of interest in this model are Φ_a, Φ, Ψ_1 and Ψ_2.

The EM Algorithm was developed for this purpose and run on a trial set of data. The results were similar to those obtained in the old model (See Table 12).

Three issues which the users of the EM Algorithm must contend with are the criterion (1), non-convergence (2) and restriction problems (3).

TABLE 12

Summary Statistics of the Four Parameter Latent Model for Balanced and Unbalanced Data Sets with 100 Groups

Ph Matrix

		Expected	Balance (h)	Unbalance (k)
Ph(1)	Mean	64.000	64.040	64.280
	SD		3.710	3.470
	MSE		13.766	12.128
	Bias(B)		0.040	0.280
	B/E		0.001	0.004
	Ratio 1			-0.119
Ph(2)	Mean	8.000	7.160	7.130
	SD		3.990	4.780
	MSE		16.704	23.689
	Bias(B)		-0.840	-0.870
	B/E		-0.105	-0.109
	Ratio 1			0.418
Ph(3)	Mean	5.000	4.380	4.340
	SD		1.600	1.480
	MSE		2.987	2.674
	Bias(B)		-0.620	-0.660
	B/E		-0.124	-0.132
	Ratio 1			-0.105
Ph(4)	Mean	40.000	39.430	39.720
	SD		10.440	10.860
	MSE		109.355	118.027
	Bias(B)		-0.570	-0.280
	B/E		-0.014	-0.007
	Ratio 1			0.079
Ph(5)	Mean	7.000	7.040	7.000
	SD		6.880	6.900
	MSE		47.336	47.610
	Bias(B)		0.040	0.000
	B/E		0.006	0.000
	Ratio 1			0.006
Ph(6)	Mean	107.000	102.600	100.940
	SD		15.970	18.540
	MSE		276.552	384.536
	Bias(B)		-4.400	-6.060
	B/E		-0.041	-0.057
	Ratio 1			0.390

TABLE 12(Continued)

Summary Statistics - Ph Matrix

	Balance	Unbalance
Mean of B/E	-0.046	-0.050
SD of B/E	0.056	0.059
Mean of Ratio 1		0.112
SD of Ratio 1		0.238

Om Matrix

		Expected	Balance	Unbalance
Om(1)	Mean	25.000	24.490	24.420
	SD		1.020	0.900
	MSE		1.329	1.184
	Bias(B)		-0.510	-0.580
	B/E		-0.020	-0.023
	Ratio 1			-0.110
Om(2)	Mean	10.000	11.780	11.670
	SD		1.360	1.410
	MSE		5.370	5.087
	Bias(B)		1.780	1.670
	B/E		0.178	0.167
	Ratio 1			-0.053
Om(3)	Mean	20.000	18.270	18.360
	SD		1.080	1.510
	MSE		4.492	5.269
	Bias(B)		-1.730	-1.640
	B/E		-0.087	-0.082
	Ratio 1			0.173
Om(4)	Mean	15.000	16.920	16.940
	SD		1.540	1.370
	MSE		6.468	6.059
	Bias(B)		1.920	1.940
	B/E		0.128	0.129
	Ratio 1			-0.063
Om(5)	Mean	10.000	12.950	13.220
	SD		4.200	4.060
	MSE		27.309	28.004
	Bias(B)		2.950	3.220
	B/E		0.295	0.322
	Ratio 1			0.025

TABLE 12(Continued)

Om Matrix

		Expected	Balance	Unbalance
Om(6)	Mean	35.000	32.320	32.510
	SD		2.710	3.380
	MSE		15.325	18.313
	Bias(B)		-2.680	-2.490
	B/E		-0.077	-0.071
	Ratio 1			0.195

Summary Statistics - Om Matrix

Mean of B/E	0.070	0.074
SD of B/E	0.155	0.160
Mean of Ratio 1		0.028
SD of Ratio 1		0.129

Psl Matrix

			Balance	Unbalance
Psl(1)	Mean	5.000	6.710	6.700
	SD		10.380	10.270
	MSE		110.993	108.684
	Bias(B)		1.710	1.700
	B/E		0.342	0.340
	Ratio 1			-0.021
Psl(2)	Mean	6.000	15.380	15.670
	SD		6.740	6.180
	MSE		143.188	142.091
	Bias(B)		9.380	9.670
	B/E		1.563	1.612
	Ratio 1			-0.008
Psl(3)	Mean	11.000	25.330	26.330
	SD		8.410	8.910
	MSE		298.894	340.509
	Bias(B)		14.330	15.330
	B/E		1.303	1.394
	Ratio 1			0.139
Psl(4)	Mean	12.000	23.910	22.970
	SD		7.540	6.250
	MSE		214.461	172.775
	Bias(B)		11.910	10.970
	B/E		0.993	0.914
	Ratio 1			-0.194

TABLE 12(Continued)

Summary Statistics - Psl Matrix

	Balance	Unbalance
Mean of B/E	0.700	0.710
SD of B/E	0.527	0.564
Mean of Ratio 1		-0.021
SD of Ratio 1		0.137

Ps2 Matrix

		Expected	Balance	Unbalance
Ps2(1)	Mean	7.000	5.160	5.330
	SD		8.170	8.170
	MSE		70.511	69.848
	Bias(B)		-1.840	-1.670
	B/E		-0.263	-0.239
	Ratio 1			-0.009
Ps2(2)	Mean	6.000	7.130	7.370
	SD		5.700	5.620
	MSE		33.909	33.670
	Bias(B)		1.130	1.370
	B/E		0.188	0.228
	Ratio 1			-0.007
Ps2(3)	Mean	10.000	10.490	11.090
	SD		7.470	7.860
	MSE		56.068	63.100
	Bias(B)		0.490	1.090
	B/E		0.049	0.109
	Ratio 1			0.125
Ps2(4)	Mean	11.000	10.700	9.760
	SD		6.380	5.620
	MSE		40.804	33.293
	Bias(B)		-0.300	-1.240
	B/E		-0.027	-0.113
	Ratio 1			-0.184

Summary Statistics - Ps2 Matrix

	Balance	Unbalance
Mean of B/E	-0.009	-0.002
SD of B/E	0.189	0.211
Mean of Ratio 1		-0.019
SD of Ratio 1		0.127

The present study used the absolute change of the estimates as the convergence criteria. Raudenbuash (1986) used the change in the likelihood for their criteria while the gradient of the likelihood has also been suggested (See Wu (1987)). However, each has its problems. The first criteria may not reach the maximum likelihood solution since the starting point definitely affects the finishing point in such a situation. Using the likelihood or its gradient can fail if there is a chance that the matrix being estimated is singular. The algorithm, using estimate differences, will converge for a singular matrix.

The pattern of the estimation in the two sets of data which failed to converge exposes a problem. The data first converged toward the correct values then diverged. The slight change of one value in the starting matrices caused these data sets to converge. There are some articles written on the convergence of the E-M algorithm (most notably Wu (1983)) but these are for univariate cases. The multivariate case becomes much more complex. Being a linear method, the EM Algorithm goes on a slow line toward a convergence point. It is much more susceptible to any local maxima or minima than Raphson-Newton or any other quadratic procedure which may jump them on its journey toward the maximum likelihood estimate.

The third issue involves the restriction of the model. The less restrictions placed on the model the better the estimation appears to be. If the model is wrongly restricted, however, the EM Algorithm will still converge yielding bad (but attractive) results with no indication that a problem exists. It becomes imperative that the Maximum Likelihood Ratio test or a similar test be used to test the fit of the model.

2. Future Exploration.

There is no clear choice as the best criterion for this method. This convergence criterion for this algorithm used the absolute change of the estimates. Alternative choices for the convergence criterion are the change in the model's likelihood or the likelihood's gradient. Each has its advantages. Since the final criteria used in this study was affected by the starting values of the estimated matrices, the other choices of criterion might yield closer consistent estimates. However if any of the matrices is singular, the likelihood will approach infinity and likely fail to converge.

Future models can be expanded to larger more restrictive and complicated models. The only problem facing these models is the number of iterations necassary for convergence. As the models become more complicated, more iterations are required to reach convergence. New developments arising in the work on the E-M algorithm might shorten this process. The E-M algorithm however must be derived separately for each model to which it is applied limiting the generalization from one model to another.

Another factor not examined here but of importance is the unbalancedness of the sample. The degree to which the data contains unbalanced groups may or may not affect the estimation procedure. Using the unbalanced design a lower and upper limit of the MSE could be found for the sample. Literature indicates that the relative size of the matrices of the random model can affect the reliability of the estimates.

Lastly the expansion of this model into the unbalanced design is important for educational research. This procedure opens the way for more complicated multilevel analysis such as the causal modeling of Joreskog.

APPENDICES

APPENDIX A

EQUATIONS FOR THE ESTIMATION OF THE COVARIANCE COMPONENTS OF THE TWO-PARAMETER MODEL USING THE EM ALGORITHM

The model of the two-parameter (unrestricted) model is:

$$\underline{Y}_{ij} = \underline{\mu} + \underline{\gamma}_i + \underline{\varepsilon}_{ij}$$

$\underline{Y}$ is a p x 1 vector of observed data.

$\underline{\mu}$ is a p x 1 vector of grand means.

$\underline{\gamma}$ is a p x 1 vector of group effects.

$\underline{\varepsilon}$ is a p x 1 vector of individual effects.

The EM algorithm is used to estimate the two covariance matrices of this model, Σ_γ and Σ_E. Both matrices are assumed to come from multivariate normal distributions:

$$\underline{\gamma} \sim N(\underline{0} , \Sigma_\gamma)$$
$$\underline{\varepsilon} \sim N(\underline{0} , \Sigma_E) .$$

E-Step

The conditional expectations of γ and E are:

$$E\begin{pmatrix}\mu|Y\\ \gamma|Y\end{pmatrix}-\begin{pmatrix}\mu^*\\ \gamma^*\end{pmatrix}=\begin{matrix}\Sigma_{i=1}^{k} W Q_i \bar{Y}\\ 1_k \otimes \Sigma_\gamma Q_i(\bar{Y}_i - \mu^*)\end{matrix}$$

where $Q_i = (\Sigma_\gamma + (1/n_i)\Sigma_\epsilon)^{-1}$

$W = [\Sigma_{i=1}^{k} Q_i]^{-1}$

M-Step

The maximum likelihood equations of the parameters for the M-step are:

$$\hat{\Sigma}_\gamma = \Sigma_\gamma - \frac{1}{k}\sum_{i=1}^{k}[\Sigma_\gamma(Q_i - Q_i(A - W)Q_i)\Sigma_\gamma]$$

where $A = (\bar{Y}_i - \mu^*)(\bar{Y}_i - \mu^*)'$

$$\hat{\Sigma}_\epsilon = \Sigma_\epsilon - [(1/N)\sum_{i=1}^{k}(1/n_i))\ \Sigma_\epsilon[Q_i - Q_i(A - W)Q_i]\Sigma_\epsilon - B - \Sigma_\epsilon$$

where $B = (\underline{Y} - \lambda_a\underline{\theta}^* - \underline{\mu}^*)(\underline{Y} - \lambda_a\underline{\theta}^* - \underline{\mu}^*)'$

APPENDIX B

EQUATIONS FOR THE ESTIMATION OF THE COVARIANCE COMPONENTS OF THE FOUR-PARAMETER MODEL USING THE EM ALGORITHM

The four-parameter (unrestricted) model is:

$$\underline{Y}_{ij} = \underline{\mu} + \lambda_a \underline{\theta}_i + \lambda \underline{\alpha}_{ij} + \underline{\epsilon 1}_{ij} + \underline{\epsilon 2}_{ij}$$

$\underline{Y}$ is a p x 1 vector of observed data.

$\underline{\mu}$ is a p x 1 vector of grand means.

λ_a is a p x q matrix connecting the p observed variables with q latent group variables.

$\underline{\theta}$ is a q x 1 vector ($q \leq p$) of the latent group effects.

λ is a p x r matrix connecting the p observed variables with r latent group variables.

$\underline{\alpha}$ is a r x 1 vector ($r \leq p$) of the latent individual effects.

$\underline{\epsilon 1}$ is a p x 1 vector of the latent individual errors.

$\underline{\epsilon 2}$ is a p x 1 vector of the latent group errors.

The EM algorithm is used to estimate the four covariance matrices of this model, Φ_a, Φ, Ψ_1 and Ψ_2. The matrices are assumed to come from multivariate normal distributions:

$$\underline{\theta} \sim N(\underline{0}, \Phi_a) \qquad \underline{\alpha} \sim N(\underline{0}, \Phi)$$

$$\underline{\epsilon 1} \sim N(\underline{0}, \Psi_1) \qquad \underline{\epsilon 2} \sim N(\underline{0}, \Psi_2).$$

E-Step

The conditional expectations of $\underline{\theta}$, $\underline{\alpha}$, $\underline{\epsilon 1}$ and $\underline{\epsilon 2}$ are:

$$(5.18)\quad E\begin{pmatrix} \underline{\mu}|Y \\ \underline{\theta}|Y \\ \underline{\epsilon 2}|Y \\ \underline{\alpha}|Y \end{pmatrix} - \begin{pmatrix} \underline{\mu}^* \\ \underline{\theta}^* \\ \underline{\epsilon 2}^* \\ \underline{\alpha}^* \end{pmatrix} = \begin{matrix} \Sigma_{i=1}^{k} W\, Q_i \bar{Y} \\ 1_k \otimes \Phi_a \lambda_a' Q_i (\bar{Y}_i - \underline{\mu}^*) \\ 1_k \otimes \Psi_2 Q_i (\bar{Y}_i - \underline{\mu}^*) \\ 1_N \otimes \Phi\, \lambda' M (Y_{ij} - \lambda_a \theta^* - \underline{\mu}^* - \underline{\epsilon 2}^*) \end{matrix}$$

where $M = (\lambda\, \Phi\, \lambda' + \Psi_1)^{-1}$

$Q_i = (\lambda_a \Phi_a \lambda_a' + \Psi_2 + (1/n_i) M^{-1})^{-1}$

$W = [\Sigma_{i=1}^{k} Q_i]^{-1}$

M-Step

The equation can be clarified by substituting By substituting (5.18) for and (5.12) for $D(\theta^*)$. The maximum likelihood estimate becomes:

$$\hat{\Phi}_a = \Phi_a - \frac{1}{k} \sum_{i=1}^{k} [\Phi_a \lambda_a' (Q_i - Q_i\, (A - W)\, Q_i)]$$

$$\hat{\Psi}_2 = \Psi_2 - \frac{1}{k} \sum_{i=1}^{k} [\Psi_2 (Q_i - Q_i (A - W) Q_i) \Psi_2]$$

$$\hat{\Phi} = \Phi - \Phi\lambda [\frac{1}{N} \sum_{i=1}^{k} \Phi\lambda' ((1/n_i)[Q_i - Q_i (A - W) Q_i] - M[B - (n_i - 1) M^{-1}] M)\, \lambda\, \Phi$$

$$\hat{\Psi}_1 = \Psi_1 - k^{-1}\sum_{i=1}^{k}\Psi_1((1/n_i)[Q_i - Q_i(A - W)Q_i] - M[B-(n_i - 1)M^{-1}]M)\Psi_1$$

where $A = (\bar{Y}_i - \mu^*)(\bar{Y}_i - \mu^*)'$

$B = (Y_{ij} - \lambda\,\theta^* - \mu^*)(Y - \lambda_a\underline{\theta}^* - \underline{\mu}^*)'$

APPENDIX C

COMPUTER PROGRAM FOR THE ESTIMATION OF THE THREE PARAMETER LATENT MODEL IN SAS

```
*                        SECTION 1 - PART 1                                ;
*                                                                          ;
* THIS SECTION CREATES THE SAMPLE DATASET FOR USE IN THE E-M ALGORITHM      ;
* STEPS.  EACH DATA POINT CONSISTS OF THREE COMPONENTS, LATENT WITHIN       ;
* (OM), LATENT BETWEEN (PH) AND ERROR (PS).  THE OBJECT IS TO USE          ;
* PATTERN MATRICES L AND LA TO CONVERT THE  3 X 3 LATENT MATRICES INTO     ;
* 4 X 4 MATRICES OF OBSERVED VALUES.  THE ERROR MATRIX IS ALWAYS A         ;
* 4 X 4 MATRIX OF MEASUREMENT ERRORS OF THE OBSERVED VALUES.               ;
*                                                                          ;
* 1. SEED IS ANY RANDOM NUMBER USED TO CREATE RANDOM VALUES FROM A         ;
*    RANDOM GENERATOR (NORMAL).                                            ;
*                        USED IN STUDY                                     ;
*    25 GROUPS                                  100 GROUPS                 ;
*      10199                                                               ;
*      50199                                                               ;
*      80199                                                               ;
*     100199                                                               ;
*     110199                                                               ;
*                                                                          ;
* 2. CIRCLE IS A COUNTER USED TO LOOP THROUGH THE PROGRAM CREATING         ;
*    DIFFERENT DATA SETS FOR ANALYSIS.                                     ;
* 3. PAT IS A Z X 2 MATRIX OF THE NUMBER OF STUDENTS IN THE                ;
*    GROUPS.  NO1 HAS THE NUMBER OF SUBJECTS IN GROUPS - NO2 HAS THE       ;
*    NO OF GROUPS OF THAT SIZE.                                            ;
*                                                                          ;
*    FOR UNBALANCED 25 GROUPS:            |  FOR BALANCED 25 GROUPS:      ;
*     PAT=10 5/20 5/30 5/40 5/50 5;       |   PAT=30 25;                   ;
*    FOR UNBALANCED 100 GROUPS:           |  FOR BALANCED 100 GROUPS:     ;
*     PAT=10 20/20 20/30 20/40 20/50 20;  |   PAT=30 100;                  ;
*                                                                          ;
* 4. OM IS THE PARAMETER OF THE WITHIN COVARIANCE MATRIX  OF THE           ;
*    POPULATION.                                                           ;
* 5. PH IS THE PARAMETER OF THE BETWEEN COVARIANCE MATRIX OF THE           ;
*    POPULATION.                                                           ;
* 6. PS IS THE PARAMETER OF THE ERROR COVARIANCE MATRIX OF THE             ;
*    POPULATION.                                                           ;
* 7. L IS A PATTERN MATRIX CREATING LINEAR COMBINATIONS OF THE LATENT      ;
*    VARIABLES IN PH.                                                      ;
* 8. LA IS A PATTERN MATRIX CREATING LINEAR COMBINATIONS OF THE LATENT     ;
*    VARIABLES IN OM.                                                      ;
*                                                                          ;
*
PROC MATRIX ;
SEED=10199;
CIRCLE=0;
```

```
PAT=30 25;
OM = 25 10 15/ 10 20 10/ 15 10 35;
PH = 64  8 40/  8  5  7/ 40  7 107;
PS =  5  0  0  0/  0  6  0  0/  0  0  11 0/ 0  0  0  12;
L = 1   0.5   0.5/
    1   0.5  -0.5/
    1  -0.5   0.5/
    1  -0.5  -0.5;
LA= 1   0.5   0.5/
    1   0.5  -0.5/
    1  -0.5   0.5/
    1  -0.5  -0.5;
NOTE 'THESE ARE THE PARAMETER VALUES AND PATTERN OF SIZES'  ;
PRINT PAT WITHIN BETWN ERR;
*                                                                    ;
*                        SECTION 1 - PART 2                          ;
*                                                                    ;
* THREE DIFFERENT VECTORS OF DATA ARE NEEDED, ONE FOR PH, ONE FOR OM  ;
* AND ONE FOR PS.  THESE ARE INDEPENDENT RANDOM VARIABLES.           ;
* FOR LATER USE THE CHOLESKYS OF OUR PARAMETER MATRICES ARE NEEDED.  ;
*                                                                    ;
* 9. CHOLOM IS THE CHOLESKY OF OM.                                   ;
*10. CHOLPH IS THE CHOLESKY OF PH.                                   ;
*11. CHOLPS IS THE CHOLESKY OF PS.                                   ;
*12. A IS VECTOR OF 21330 VALUES GENERATED AT RANDOM FROM A POPULATION ;
*    OF VALUES WITH A MEAN OF 0 AND A VARIANCE OF 1 FROM SAS         ;
*    SUBROUTINE NORMAL.                                              ;
*13. Z IS A VECTOR OF 3000 VALUES EQUAL TO THE INDIVIDUAL VALUES FOR ;
*      OM.                                                           ;
*14. Z1 IS A VECTOR OF 100 VALUES EQUAL TO THE GROUP VALUES FOR PH.  ;
*15. Z2 IS A VECTOR OF 3000 VALUES EQUAL TO THE INDIVIDUAL VALUES FOR ;
*      PS.                                                           ;
*16. C, C1 AND C2 ARE THE VARIANCE-COVARIANCES FOR THE THREE Z VECTORS.;
*    THESE MATRICES SHOULD BE IDENTITY MATRICES.                     ;
*                                                                    ;
CHOLOM = HALF(OM);
CHOLPH = HALF(PH);
CHOLPS = HALF(PS);
BEGIN:  CIRCLE=CIRCLE+1;
A = J.(21300,1,0);
I = 1;
L: A(I,1)=NORMAL(SEED);
I=I+1;
IF I<= 21300 THEN GO TO L;
Z = a(1:3000,1)||a(3001:6000,1)||a(6001:9000,1);
Z1= a(21001:21100,1)||a(21101:21200,1)||a(21201:21300,1);
Z2=a(9001:12000,1)||a(12001:15000,1)||
                              a(15001:18000,1)||a(18001:21000,1)    ;
TOTMI=NROW(Z)-1                                                     ;
TOTMIG=NROW(Z1)-1                                                   ;
C = (Z'*Z)#/TOTMI                                                   ;
C1= (Z1'*Z1)#/TOTMIG                                                ;
C2= (Z2'*Z2)#/TOTMI                                                 ;
```

```
NOTE 'THESE ARE THE VAR-COV OF THE RANDOM DATA (NO TRANS)' ;
PRINT C C1 C2;
*                                                                       ;
*                       SECTION 1 - PART 3                              ;
*                                                                       ;
* BY MULTIPLYING RANDOM DATA FROM A POPULATION WITH MEAN 0 AND VARIANCE;
*    OF 1 BY THE CHOLESKY OF A MATRIX, A VECTOR IS CREATED WHICH WILL  ;
*    RECREATE THAT MATRIX.                                             ;
*                                                                       ;
*17. Y IS THE PRODUCT OF Z AND CHOLW.                                  ;
*18. Y1 IS THE PRODUCT OF Z1 AND CHOLB.                                ;
*19. Y2 IS THE PRODUCT OF Z2 AND CHOLERR.                              ;
*20. D, D1 AND D2 ARE THE VARIANCE-COVARIANCES FOR THE THREE Y VECTORS-;
*      THESE MATRICES SHOULD BE CLOSE TO THE PARAMETER MATRICES.        ;
*                                                                       ;
Y = Z  * CHOLW;
D = (Y'*Y)#/TOTMI;
Y1= Z1 * CHOLB;
D1 = (Y1'*Y1)#/TOTMIG;
Y2= Z2 * CHOLERR;
D2=  (Y2'*Y2)#/TOTMI;
NOTE 'THESE ARE THE VAR-COV OF THE TRANSFORMED DATA'      ;
PRINT D D1 D2 ;
*                                                                       ;
*                       SECTION 1 - PART 4                              ;
*                                                                       ;
* BY MULTIPLYING VECTORS Z AND Z1 TO L AND LA, THE OBSERVED VALUES FOR ;
*    FOR EACH INDIVIDUAL ARE CREATED. INSTEAD OF THREE MEASURES PER     ;
*    INDIVIDUAL THERE WILL BE FOUR.  (THE ERROR MATRIX WAS CREATED IN   ;
*    TERMS OF ERRORS FOR EACH OBSERVED VARIABLES AND IS ALREADY 4 X 4.);
*                                                                       ;
*21. X IS THE PRODUCT OF Y AND L.                                      ;
*22. X1 IS THE PRODUCT OF Y1 AND LA.                                   ;
*                                                                       ;
X = Y * L'   ;
X1= Y1 * LA'  ;
*                                                                       ;
*                       SECTION 1 - PART 5                              ;
*                                                                       ;
* BY ADDING VECTORS X, YY1 AND Y2 TOGETHER, A TOTAL SCORE IS ACHIEVED  ;
*    FOR EACH INDIVIDUAL.  THESE SCORES OBVIOUSLY CONTAIN THE THREE     ;
*    VARIANCE COMPONENTS.  ALL 30 INDIVIDUALS IN EACH GROUP RECEIVE     ;
*    THE SAME GROUP VECTOR(X1).  OTHERWISE EACH RECEIVES A DIFFERENT    ;
*    VALUE FROM BOTH X AND Y2.                                          ;
*                                                                       ;
*23. YY2 BECOMES A 3000 x 4 VECTOR WHICH REPEATS THE SAME X1 VALUE FOR ;
*       N(I) TIMES FOR EACH GROUP.                                      ;
*24. X BECOMES THE SUM OF X AND YY1 AND Y2.                            ;
*25. FIN IS THE VARIANCE-COVARIANCE MATRIX FOR THE FINAL SET OF DATA-  ;
*      ITS A 4 X 4 MATRIX BASED ON 3000 OBSERVATIONS.                   ;
*26. NK IS A VECTOR OF SIZE K CONTAINING THE GROUP SIZE FOR EACH GROUP.;
*27. RD REPLACES X AS THE MATRIX OF DATA.  THIS IS USED IS OTHER       ;
*      SECTIONS.                                                        ;
*                                                                       ;
```

```
MM=0;
II=1;
NN=1;
JJ: MM=MM+1;
CC=J.(PAT(II,1),1,1);
DD=(CC @ X1(NN,))  ;
YY1=YY1//DD          ;
NN=NN+1             ;
NK=NK//PAT(II,1)    ;
IF MM LT PAT(II,2) THEN GO TO JJ;
MM=0;                    II=II+1  ;
IF NN LT PAT(+,2) THEN GO TO JJ ;
F1=(YY1'*YY1)#/NROW(YY1)            ;
NOTE 'THIS IS THE VAR-COV MATRIX OF GROUP DATA FOR ALL IND'    ;
PRINT F1       ;
RD=X+YY1+Y2                         ;
FIN= (RD'*RD)#/TOTMI                ;
NOTE 'THIS IS THE VAR-COV MATRIX OF THE DATA TO BE USED'       ;
PRINT FIN      ;
FREE MM NN II X X1 Y Y1 Y2 D D1 D2 E E1 FIN I YY1 Z Z1 Z2 CC DD ;
FREE F1 C C1 C2 A EE EE1 EE2 TOTE WITHIN BETWN ERR TOTMI TOTMIG ;
*                                                                  ;
*                                                                  ;
*                       END OF SECTION 1                           ;
*                                                                  ;
*                                                                  ;
* AT THIS POINT IT BECOMES IMPORTANT TO REALIZE THAT ALL THE LINES  ;
* ABOVE DEAL ONLY WITH CREATING THE DATA FOR THIS ANALYSIS.  THEY   ;
* CAN BE DROPPED IN USING THE EM ALGORITHM.  TO USE THE REST OF THE ;
* PROGRAM WITHOUT THE PRIOR LINES, THE FOLLOWING LINES MUST BE PLACED ;
* AT THE TOP OF THE PROGRAM (REMOVING THE * FROM THE FRONT - SEE SAS ;
* FOR THE FETCH COMMAND) :                                         ;
*                                                                  ;
*PROC MATRIX                                                       ;
*FETCH RD                                                         ;
*FETCH LA                                                         ;
*FETCH L                                                          ;
*FETCH NK                                                         ;
*                                                                  ;
*                                                                  ;
*                       SECTION 2 - PART 1                         ;
*                                                                  ;
* THIS SECTION USES THE EM ALGORITHM TO GET ESTIMATES OF THE       ;
* UNRESTRICTED MODEL.  THE BETWEEN AND WITHIN VARIANCE-COVARIANCE  ;
* MATRICES ARE ESTIMATED WITH NO STRUCTURE APPLIED.  THIS FIRST PART ;
* TURNS OUT THE SUFFICIENT STATISTICS FOR THE SAMPLE DATA NEEDED   ;
* IN PART 2 AND IN PART 3.                                         ;
*                                                                  ;
* 1. K IS THE NUMBER OF GROUPS IN THE SAMPLE.                      ;
* 2. P IS THE NUMBER OF OBSERVED VARIABLES IN THE SAMPLE.          ;
* 3. N IS THE TOTAL NUMBER OF OBSERVATIONS IN THE SAMPLE.          ;
* 4. YM IS A KP VECTOR OF THE GROUP MEANS.                         ;
* 5. SS IS A KP X KP MATRIX OF EACH GROUP'S SUM OF SQUARE/NK.      ;
```

```
ZERO1=J.(NROW(L),NROW(L),0)          ;
ZERO2=J.(NCOL(L),NCOL(L),0)          ;
ZERO3=J.(NCOL(LA),NCOL(LA),0)        ;
K=NROW(NK)                ; *NO OF CLASSES          -    K          ;
P=NCOL(RD)                ; *NO OF OBSERVED VARIABLES - P          ;
N=NK(+,)                  ; *NO OF TOTAL INDIVIDUALS  - N          ;
GRP=J.(NROW(NK),1,1)      ; *VECTOR OF 1'S, K X 1                  ;
D=0                       ;
B1=0                      ;
DO I=1 TO K                                                ;
C =D+1                                                     ;
D =C+NK(I,)-1                                              ;
A =RD(C:D,)                                                ;
B =A(+,)*(1#/NK(I,))                                       ;
YM =YM//B'                ; *GROUP MEANS                   ;
E =(A'*A)*(1#/NK(I,))                                      ;
SS =SS//E                 ; *SS FOR EACH GROUP             ;
B1=((B'*B)*NK(I,))+B1     ; *SS/K OF THE GROUP MEANS       ;
END                                                        ;
*                                                                        ;
*                      SECTION 2 - PART 2                                ;
*                                                                        ;
* STARTING VALUES ARE NEEDED FOR THE BETWEEN GROUP COVARIANCE MATRIX     ;
* PHI AND THE WITHIN COVARIANCE MATRIX PSI.  THE MLE FOR EQUAL N'S       ;
* WILL BE USED WITH NK (NUMBER OF STUDENTS IN A GROUP) REPLACED BY THE   ;
* HARMONIC MEAN OF NK.                                                   ;
*                                                                        ;
* 6. NH IS THE HARMONIC NK OF THE GROUPS.                                ;
* 7. PHI IS THE BETWEEN GROUPS COVARIANCE MATRIX.                        ;
* 8. PSI IS THE WITHIN GROUPS COVARIANCE MATRIX.                         ;
*                                                                        ;
NH=(1#/SUM(INV(DIAG(NK))))*K                                             ;
PSI=((RD'*RD)-B1)*1#/(N-K)                                               ;
PHI=(1#/NH)*((B1-(RD(+,)'*RD(+,)*(1#/N)))*1#/(K-1)-PSI)                  ;
NOTE 'HERE ARE THE STARTING MATRICES'                                    ;
PRINT PHI PSI                                                            ;
FREE  B1 B E NH A                                          ;
*                                                                        ;
*                      SECTION 2 - PART 3                                ;
*                                                                        ;
* THIS PART CREATES THE CONDITIONAL VALUES FOR THE MEAN AND GROUP        ;
* EFFECT FOR PHI AND PSI ESTIMATES.  THERE ARE FOUR IMPORTANT VARIABLES; ;
* CREATED HERE.   THEY ARE CREATED IN SUBROUTINES ALPHAB AND ALPHAU.     ;
* THIS IS PART OF THE INTERATIVE LOOP, THE E STEP.                       ;
*                                                                        ;
*  9. U IS A P X 1 VECTOR OF CONDITIONAL MEANS.                          ;
* 10. TH IS A K X P MATRIX OF GROUP EFFECTS.                             ;
* 11. Q IS A KP X P WEIGHTING FACTORS FOR THE GROUPS.                    ;
* 12. W IS A P X P WEIGHTING FACTOR CALCULATED FROM THE Q'S.             ;
*                                                                        ;
BUDDY=0                                                    ;
BUD:BUDDY=BUDDY+1                                          ;
IF NROW(PAT) EQ 1 THEN LINK ALPHAB                         ;
   ELSE LINK ALPHAU                                        ;
```

```
DO I=1 TO K                                                    ;
C =(P*I)-P+1                                                   ;
D =P*I                                                         ;
A =PHI*Q(C:D,)*(YM(C:D,)-U)                                    ;
TH=TH//A                                                       ;
END                                                            ;
FREE A                                                         ;
*                                                                                ;
*                       SECTION 2 - PART 4                                       ;
*                                                                                ;
* THIS PART CALCULATES THE MAXIMUM LIKELIHOOD VALUES FOR PHI AND PSI             ;
* USING THE DATA AND THE CONDITIONAL VARIABLES. (M-STEP). THIS PROGRAM           ;
* WILL KEEP LOOPING TO THE LAST PART    UNTIL THE DIFFERENCES IN PHI             ;
* AND PSI, AND THE NEW ESTIMATES OF PHI AND PSI ARE LESS THAN .01.               ;
*                                                                                ;
* 13. ONE1 IS THE DIFFERNECE BETWEEN PHI ON THE LAST ITERATION AND               ;
*        THE NEW ESTIMATES OF PHI.                                               ;
* 14. TWO1 IS THE DIFFERNECE BETWEEN PSI ON THE LAST ITERATION AND               ;
*        THE NEW ESTIMATES OF PSI.                                               ;
*                                                                                ;
E=J(P,P,0)                                                     ;
F=J(P,P,0)                                                     ;
DO I=1 TO K                                                    ;
C =(P*I)-P+1                                                   ;
D =P*I                                                         ;
A =TH(C:D,)+U                                                  ;
E =E+NK(I,)*((SS(C:D,))-(YM(C:D,)*A')
      -(A*YM(C:D,)')+(A*A')+(PHI*Q(C:D,)*
      (PSI*(1#/NK(I,))+(W*Q(C:D,)*PHI-2*W))))                          ;
F =F+Q(C:D,)-(Q(C:D,)*((YM(C:D,)-U)*(YM(C:D,)-U)'+W)*Q(C:D,))      ;
END                                                            ;
FREE A                                                         ;
PH1 =PHI-(PHI*((1#/K)*F)*PHI)                                  ;
ONE1=PH1-PHI                                                   ;
PS1 =((1#/N)*E)+W                                              ;
TWO1=PS1-PSI                                                   ;
PH1D=DIAG(PH1)                                                 ;
PS1D=DIAG(PS1)                                                 ;
PH1D=PH1D<>ZERO1                                               ;
PS1D=PS1D<>ZERO1                                               ;
PH1 = PH1-DIAG(PH1)+PH1D                                       ;
PS1 = PS1-DIAG(PS1)+PS1D                                       ;
PHI=PH1                                                        ;
PSI=PS1                                                        ;
FREE PS1 Q TH PH1 PH1D PS1D                                    ;
IF BUDDY GT 250 THEN GO TO FINAL;
IF MAX(ABS(ONE1)) LT 0.01 AND MAX(ABS(TWO1)) LT 0.01
   THEN GO TO FINAL                                            ;
   ELSE GO TO BUD                                              ;
FINAL:PRINT  BUDDY PHI PSI ONE1 TWO1  U                        ;
FREE PS1 ONE1 TWO1 U BUDDY                                     ;
```

```
*                                                                    ;
*                                                                    ;
* END OF SECTION 2                                                   ;
*                                                                    ;
*                        SECTION 3 - PART 1                          ;
*                                                                    ;
* THIS SECTION USES THE E-M ALGORITHM TO GET ESTIMATES OF THE        ;
* RESTRICTED MODEL.  PH, OM AND PS ARE ESTIMATED WITH STRUCTURE      ;
* APPLIED TO THE MODEL.  THIS FIRST PART MAKES USE OF PHI AND PSI    ;
* FROM THE LAST SECTION TO GET OPENING ESTIMATES OF PH, OM AND PSI.  ;
*                                                                    ;
* 1. PH IS THE LATENT GROUP LEVEL VAR-COV.                           ;
* 2. OM IS THE LATENT IND LEVEL VAR-COV.                             ;
* 3. PS IS THE OBSERVED VARIABLES ERROR MATRIX.                      ;
* 4. S IS THE DIMENSION OF PH.                                       ;
* 5. R IS THE DIMENSION OF OM.                                       ;
*                                                                    ;
Y1=INV(L'*L)                                                         ;
Y2=INV(LA'*LA)                                                       ;
PH=Y1*L'*PHI*L*Y1                                                    ;
OM=Y2*LA'*PSI*LA*Y2                                                  ;
PS=PSI+PHI-L*PH*L'-LA*OM*LA'                                         ;
NOTE 'THESE ARE THE STARTING VALUES IN THIS STEP'                    ;
PRINT PH OM PS                                                       ;
S=NCOL(L)                ; *NO OF LATENT CLASS VARIABLES - S         ;
R=NCOL(LA)               ; *NO OF LATENT IND VARIABLES   -R          ;
*                                                                    ;
*                        SECTION 3 - PART 2                          ;
*                                                                    ;
* THIS PART CREATES THE CONDITIONAL VALUES FOR THE MEAN AND GROUP    ;
* EFFECT FOR PH, OM, PS  ESTIMATES.  THERE ARE FOUR IMPORTANT VARIABLES;
* CREATED HERE.   THEY ARE CREATED IN SUBROUTINES BETAB AND BETAU.   ;
* THIS IS PART OF THE INTERATIVE LOOP, THE E STEP.                   ;
*                                                                    ;
*  6. U IS A P X 1 VECTOR OF CONDITIONAL MEANS.                      ;
*  7. TH IS A K X P MATRIX OF GROUP EFFECTS.                         ;
*  8. Q IS A KP X P WEIGHTING FACTORS FOR THE GROUPS.                ;
*  9. W IS A P X P WEIGHTING FACTOR CALCULATED FROM THE Q'S.         ;
* 10. MM IS A VAR-COV OF THE IND LEVEL MATRICES, OM AND PS.          ;
*                                                                    ;
BUDDY1=0                  ;
BUD1: MM=(LA*OM*LA'+PS)   ;
BUDDY1=BUDDY1+1           ;
M=INV(MM)                ; *INV OF WITHIN VARIANCES       -  P X P;
IF NROW(PAT) EQ 1 THEN LINK BETAB                                    ;
   ELSE LINK BETAU       ;
DO I=1 TO K              ;
C=(P*I)-P+1              ;
D=P*I                    ;
B=PH*L'*Q(C:D,)*(YM(C:D,)-U)  ;
TH=TH//B                 ; *COND THETA                    - KP X P;
END                      ;
FREE B                   ;
```

```
*                        SECTION 3 - PART 3                                ;
*                                                                          ;
* THIS PART CALCULATES THE MAXIMUM LIKELIHOOD VALUES FOR PH,OM AND PS      ;
* USING THE DATA AND THE CONDITIONAL VARIABLES. (M-STEP). THIS PROGRAM     ;
* WILL KEEP LOOPING TO THE LAST PART    UNTIL THE DIFFERENCES IN PH         ;
* OM AND PS AND  NEW ESTIMATES OF PH, OM AN PS  ARE LESS THAN .01.         ;
*                                                                          ;
* 11. ONE IS THE DIFFERNECE BETWEEN PH ON THE LAST ITERATION AND           ;
*        THE NEW ESTIMATES OF PH.                                          ;
* 12. TWO IS THE DIFFERNECE BETWEEN OM ON THE LAST ITERATION AND           ;
*        THE NEW ESTIMATES OF OM.                                          ;
* 13. THREE IS THE DIFFERNECE BETWEEN PS ON THE LAST ITERATION AND         ;
*        THE NEW ESTIMATES OF PS.                                          ;
*                                                                          ;
VE= J(P,P,0)                ;
B = J(P,P,0)                ;
XX= J(P,P,0)                ;
HH=LA*OM*LA'                ;
II=L*PH*L'                  ;
DO I=1 TO K                 ;
C=(P*I)-P+1                 ;
D=P*I                       ;
CC=(S*I)-S+1                ;
DD=S*I                      ;
FF=L*TH(CC:DD,)+U           ;
GG=Q(C:D,)*W*Q(C:D,)        ;
KK=II+((1#/NK(I,))*HH)      ;
VE=VE+(Q(C:D,)-Q(C:D,)*( (YM(C:D,)-U) * (YM(C:D,)-U)' - W )*Q(C:D,))   ;
A=M*NK(I,)*(SS(C:D,)-(YM(C:D,)*FF')-(FF*YM(C:D,)')+(FF*FF'))*M         ;
B=B+((1#/NK(I,))*GG+M*L*PH*L'*Q(C:D,)) + A                             ;
XX=XX+(PS*A*PS+(NK(I,))*((1#/NK(I,))*HH*M*HH-2*KK*Q(C:D,)*W
    +KK*(GG-Q(C:D,))*KK))                                              ;
END                                                                    ;
ONE=-((1#/K)*(PH*L'*VE*L*PH)) ;  *EST OF PHI (GROUP LEVEL) -  S X S;
TWO=(OM*LA'*(B*(1#/N)-M)*LA*OM)                                        ;
                                 *EST OF OMEGA (IND LEVEL) -  R X R;
THREE=DIAG(W-PS*M*PS+II+(1#/N)*XX)                                     ;
PH=PH+ONE ;                      *EST OF PHI (GROUP LEVEL) -  S X S;
OM=OM+TWO                                                              ;
PS=DIAG(PS+THREE)                                                      ;
PHD=DIAG(PH)<>ZERO2                                                    ;
OMD=DIAG(OM)<>ZERO3                                                    ;
PSD=DIAG(PS)<>ZERO1                                                    ;
PH=PH-DIAG(PH)+PHD                                                     ;
OM=OM-DIAG(OM)+OMD                                                     ;
PS=PS-DIAG(PS)+PSD                                                     ;
FREE TH Q W B VE XX B II                                               ;
IF BUDDY1 GT 250 THEN GO TO FINAL1                                     ;
IF MAX(ABS(ONE)) LT 0.01 AND MAX(ABS(TWO)) LT 0.01 AND
   MAX(ABS(THREE)) LT 0.01 THEN GO TO FINAL1                           ;
   ELSE GO TO BUD1                                                     ;
FINAL1: PRINT BUDDY1 PH OM PS ONE TWO THREE SEED                       ;
PRINT U                                                    ;
```

```
FREE NK RD YM SS N K R S                                    ;
IF CIRCLE LT 2 THEN GO TO BEGIN                             ;
STOP                                                        ;
*                                                           ;
*  HERE ARE THE SUBROUTINES                                 ;
*                                                           ;
*                ALPHAU                                     ;
*                                                           ;
ALPHAU:  TOT =J(P,P,0)                                      ;
DO I=1 TO K                                                 ;
T1 =INV((PSI*(1#/NK(I,)))+PHI)                              ;
TOT =TOT+T1                                                 ;
Q =Q//T1                                                    ;
END                                                         ;
W =INV(TOT)                                                 ;
U =W*Q'*YM                                                  ;
FREE  T1 TOT                                                ;
RETURN                                                      ;
*                                                           ;
*                ALPHAB                                     ;
*                                                           ;
ALPHAB:   T1 =INV((PSI*(1#/NK(1,)))+PHI)                    ;
Q  =GRP @ T1                                                ;
W  =INV(T1*K)                                               ;
U =W*Q'*YM                                                  ;
FREE  T1                                                    ;
RETURN                                                      ;
*                                                           ;
*                BETAU                                      ;
*                                                           ;
BETAU:  W = J(P,P,0)                                        ;
DO I=1 TO K                                                 ;
A=INV( (L*PH*L') + (MM*(1#/NK(I,)))  )                      ;
Q=Q//A                    ; *MATRIX OF Q                        - KP X P;
W=W+A                     ;
END                       ;
FREE A                    ;
W=INV(W)                  ; *COND VAR FOR U                     -  P X P;
U=W*Q'*YM                 ; *COND U                             -  P X P;
RETURN                                                      ;
*                                                           ;
*                BETAB                                      ;
*                                                           ;
BETAB:  A=INV( (L*PH*L') + (MM*(1#/NK(1,)))  )              ;
Q  =GRP @ A                                                 ;
W  =INV(A*K)                                                ;
U=W*Q'*YM                 ; *COND U                             -  P X P;
RETURN                                                      ;
*                                                           ;
```

APPENDIX D

COMPUTER PROGRAM FOR THE ESTIMATION OF THE FOUR PARAMETER LATENT MODEL IN SAS

```
*                        SECTION 1 - PART 1                                ;
*                                                                          ;
* THIS SECTION CREATES THE SAMPLE DATASET FOR USE IN THE E-M ALGORITHM      ;
* STEPS.  EACH DATA POINT CONSISTS OF THREE COMPONENTS, LATENT WITHIN       ;
* (OM), LATENT BETWEEN (PH) AND ERROR (PS).  THE OBJECT IS TO USE          ;
* PATTERN MATRICES L AND LA TO CONVERT THE  3 X 3 LATENT MATRICES INTO     ;
* 4 X 4 MATRICES OF OBSERVED VALUES.  THE ERROR MATRICES ARE ALWAYS        ;
* 4 X 4 MATRICES OF MEASUREMENT ERRORS OF THE OBSERVED VALUES.             ;
*                                                                          ;
*      THE NOMENCLATURE USED IN THIS PROGRAM IS THE SAME AS THAT           ;
*       IN THE PROGRAM IN APPENDIX 3.  REFER TO APPENDIX 3 FOR             ;
*                          DEFININTIONS.                                   ;
*                                                                          ;
* 1. SEED IS ANY RANDOM NUMBER USED TO CREATE RANDOM VALUES FROM A         ;
*    RANDOM GENERATOR (NORMAL).                                            ;
*                                                                          ;
*                      USED IN STUDY                                       ;
*                                                                          ;
*   100 GROUPS - 26298, 27309, 49329, 93369, AND 181449                    ;
*                                                                          ;
* 3. PAT IS A Z X 2 MATRIX OF THE NUMBER OF STUDENTS IN THE                ;
*    GROUPS.  NO1 HAS THE NUMBER OF SUBJECTS IN GROUPS - NO2 HAS THE       ;
*    NO OF GROUPS OF THAT SIZE.                                            ;
*                                                                          ;
*    FOR UNBALANCED 100 GROUPS:          |  FOR BALANCED 100 GROUPS:       ;
*     PAT=10 20/20 20/30 20/40 20/50 20; |   PAT=30 100;                    ;
*                                                                          ;
* 4. OM IS THE PARAMETER OF THE WITHIN COVARIANCE MATRIX  OF THE           ;
*    POPULATION.                                                           ;
* 5. PH IS THE PARAMETER OF THE BETWEEN COVARIANCE MATRIX OF THE           ;
*    POPULATION.                                                           ;
* 6. PS1 IS THE PARAMETER OF THE WITHIN ERROR COVARIANCE MATRIX OF         ;
*    THE POPULATION.                                                       ;
* 6. PS2 IS THE PARAMETER OF THE BETWEEN ERROR COVARIANCE MATRIX OF        ;
*    THE POPULATION.                                                       ;
*                                                                          ;
PROC MATRIX                              ;
SEED = 101997                            ;
CIRCLE = 0                               ;
PAT = 30 100                             ;
OM = 25 10 15/ 10 20 10/ 15 10 35;
PH = 64  8 40/  8  5  7/ 40  7 107;
PS1 =  5  0  0  0/  0  6  0  0/  0  0  11 0/ 0  0  0  12;
PS2 =  7  0  0  0/  0  8  0  0/  0  0  10 0/ 0  0  0  11;
```

```
PRINT OM PH PS1 PS2                                          ;
L = 1   0.5   0.5/
    1   0.5  -0.5/
    1  -0.5   0.5/
    1  -0.5  -0.5;
LA= 1   0.5   0.5/
    1   0.5  -0.5/
    1  -0.5   0.5/
    1  -0.5  -0.5;
EOM = L*OM*L'+PS1                                       ;
EPH = LA*PH*LA'+PS2                                        ;
PRINT EOM EPH                                                 ;
*                                                                          ;
*                        SECTION 1 - PART 2                                ;
*                                                                          ;
* FOUR DIFFERENT VECTORS OF DATA ARE NEEDED, ONE FOR PH, ONE FOR OM        ;
* ONE FOR PS1 AND ONE FOR PS2. THESE ARE INDEPENDENT RANDOM VARIABLES.     ;
* FOR LATER USE THE CHOLESKYS OF OUR PARAMETER MATRICES ARE NEEDED.        ;
*                                                                          ;
CHOLOM = HALF(OM)                                             ;
CHOLPH = HALF(PH)                                              ;
CHOLPS1 = HALF(PS1)                                              ;
CHOLPS2 = HALF(PS2)                                             ;
BEGIN:  CIRCLE=CIRCLE+1                                          ;
A = J.(21700,1,0);
I = 1;
L: A(I,1)=NORMAL(SEED);
I=I+1;
IF I<= 21700 THEN GO TO L;
Z = a(1:3000,1)||a(3001:6000,1)||a(6001:9000,1)                          ;
Z1= a(21001:21100,1)||a(21101:21200,1)||a(21201:21300,1)                 ;
Z2=a(9001:12000,1)||a(12001:15000,1)
                  ||a(15001:18000,1)||a(18001:21000,1)                   ;
Z3=a(21301:21400,1)||a(21401:21500,1)
                  ||a(21501:21600,1)||a(21601:21700,1)                   ;
TOTMI=NROW(Z)-1                                                          ;
TOTMIG=NROW(Z1)-1                                                        ;
*                                                                        ;
*                        SECTION 1 - PART 3                              ;
*                                                                        ;
* BY MULTIPLYING RANDOM DATA FROM A POPULATION WITH MEAN 0 AND VARIANCE;
*    OF 1 BY THE CHOLESKY OF A MATRIX, A VECTOR IS CREATED WHICH WILL    ;
*    RECREATE THAT MATRIX.                                               ;
*                                                                        ;
Y = Z  * CHOLOM       ;
Y1= Z1 * CHOLPH       ;
Y2= Z2 * CHOLPS1      ;
Y3= Z3 * CHOLPS2      ;
*                                                                        ;
*                        SECTION 1 - PART 4                              ;
*                                                                        ;
* BY MULTIPLYING VECTORS Z AND Z1 TO L AND LA, THE OBSERVED VALUES FOR   ;
*    FOR EACH INDIVIDUAL ARE CREATED. INSTEAD OF THREE MEASURES PER      ;
```

```
*    INDIVIDUAL THERE WILL BE FOUR.  (THE ERROR MATRIX WAS CREATED IN  ;
*    TERMS OF ERRORS FOR EACH OBSERVED VARIABLES AND IS ALREADY 4 X 4.);
*                                                                      ;
 X = Y * L'           ;
 X1= Y1 * LA'         ;
 X1=X1+Y3             ;
 X2=X +Y2             ;
*                                                                      ;
*                      SECTION 1 - PART 5                              ;
*                                                                      ;
* BY ADDING VECTORS X1 AND X2  TOGETHER, A TOTAL SCORE IS ACHIEVED     ;
*    FOR EACH INDIVIDUAL.  THESE SCORES OBVIOUSLY CONTAIN THE FOUR     ;
*    VARIANCE COMPONENTS.  ALL INDIVIDUALS IN EACH GROUP RECEIVE       ;
*    THE SAME GROUP VECTOR (X1) AND A DIFFERENT VALUE FROM X2.         ;
*                                                                      ;
MM=0;
II=1;
NN=1;
JJ: MM=MM+1;
CC=J.(PAT(II,1),1,1);
DD=(CC @ X1(NN,))  ;
YY1=YY1//DD          ;
NN=NN+1            ;
NK=NK//PAT(II,1)   ;
IF MM LT PAT(II,2) THEN GO TO JJ                                      ;
MM=0;                  II=II+1                                        ;
IF NN LT PAT(+,2) THEN GO TO JJ                                       ;
FREE MM NN II X1 Y Y1         I Z Z1 Z2 CC DD                         ;
FREE A OM PH PS1 PS2 TOTMIG                                           ;
RD=X+YY1+Y2                                                           ;
FIN= (RD'*RD)#/TOTMI                                                  ;
FREE  X Y2 FIN YY1 TOTMI Z Z1 Z2 Z3                                   ;
*                                                                     ;
*                                                                      ;
*                                                                      ;
* END OF SECTION 1                                                     ;
*                                                                      ;
* AT THIS POINT IT BECOMES IMPORTANT TO REALIZE THAT ALL THE LINES     ;
* ABOVE DEAL ONLY WITH CREATING THE DATA FOR THIS ANALYSIS.  THEY      ;
* CAN BE DROPPED IN USING THE EM ALGORITHM.  TO USE THE REST OF THE    ;
* PROGRAM WITHOUT THE PRIOR LINES, THE FOLLOWING LINES MUST BE PLACED  ;
* AT THE TOP OF THE PROGRAM (REMOVING THE * FROM THE FRONT - SEE SAS   ;
* FOR THE FETCH COMMAND) :                                             ;
*                                                                      ;
*PROC MATRIX                                                           ;
*FETCH RD                                                             ;
*FETCH LA                                                             ;
*FETCH L                                                              ;
*FETCH NK                                                             ;
*                                                                      ;
*                                                                      ;
```

```
*                      SECTION 2 - PART 1                                ;
*                                                                        ;
* THIS SECTION USES THE EM ALGORITHM TO GET ESTIMATES OF THE             ;
* UNRESTRICTED MODEL.  THE BETWEEN AND WITHIN VARIANCE-COVARIANCE       ;
* MATRICES ARE ESTIMATED WITH NO STRUCTURE APPLIED.  THIS FIRST PART    ;
* TURNS OUT THE SUFFICIENT STATISTICS FOR THE SAMPLE DATA NEEDED        ;
* IN PART 2 AND IN PART 3.                                              ;
*
;ZERO1 = J.(NROW(L),NROW(L),0)                                          ;
ZERO2 = J.(NCOL(L),NCOL(L),0)                                          ;
ZERO3 = J.(NCOL(LA),NCOL(LA),0)                                        ;
ZERO1 = DIAG(ZERO1)                                                     ;
ZERO2 = DIAG(ZERO2)                                                     ;
ZERO3 = DIAG(ZERO3)                                                     ;
K=NROW(NK)               ; *NO OF CLASSES          -    K             ;
P=NCOL(RD)               ; *NO OF OBSERVED VARIABLES - P               ;
N=NK(+,)                 ; *NO OF TOTAL INDIVIDUALS  - N               ;
GRP=J.(NROW(NK),1,1)     ; *VECTOR OF 1'S, K X 1                       ;
S=NCOL(L)                ; *NO OF LATENT CLASS VARIABLES - S           ;
R=NCOL(LA)               ; *NO OF LATENT IND VARIABLES   -R            ;
D=0                                                                    ;
B =J(P,P,0)                                                            ;
B1=J(P,P,0)                                                            ;
DO I=1 TO K                                                            ;
C =D+1                                                                 ;
D =C+NK(I,)-1                                                          ;
A =RD(C:D,)                                                            ;
A2=A(+,)*(1#/NK(I,))                                                   ;
YM= YM//A2'              ; *GROUP MEANS                                ;
EE = (A'*A)                                                            ;
SS=SS//(EE*(1#/NK(I,)))                                                ;
B =B+EE-NK(I,)*A2'*A2                                                  ;
B1=((A2'*A2)*NK(I,))+B1  ; *SS/K OF THE GROUP MEANS                    ;
END                                                                    ;
*                                                                                ;
*                      SECTION 2 - PART 2                                        ;
*                                                                                ;
* STARTING VALUES ARE NEEDED FOR THE BETWEEN GROUP COVARIANCE MATRIX             ;
* PHI AND THE WITHIN COVARIANCE MATRIX PSI.  THE MLE FOR EQUAL N'S               ;
* WILL BE USED WITH NK (NUMBER OF STUDENTS IN A GROUP) REPLACED BY THE           ;
* HARMONIC MEAN OF NK.                                                           ;
*
;NH=(1#/SUM(INV(DIAG(NK))))*K                                                   ;
PSI=((RD'*RD)-B1)*1#/(N-K)                                                      ;
PHI=(1#/NH)*((B1-(RD(+,)'*RD(+,)*(1#/N)))*1#/(K-1)-PSI)                        ;
FREE     E NH A                                                                ;
*                                                                                ;
*                      SECTION 2 - PART 3                                        ;
*                                                                                ;
* THIS PART CREATES THE CONDITIONAL VALUES FOR THE MEAN AND GROUP                ;
* EFFECT FOR PHI AND PSI ESTIMATES.  THERE ARE FOUR IMPORTANT VARIABLES;
* CREATED HERE.   THEY ARE CREATED IN SUBROUTINES ALPHAB AND ALPHAU.             ;
* THIS IS PART OF THE INTERATIVE LOOP, THE E STEP.                               ;
*                                                                                ;
```

```
BUDDY=0                                                          ;
BUD:BUDDY=BUDDY+1                                                ;
IF NROW(PAT) EQ 1 THEN LINK ALPHAB                               ;
   ELSE LINK ALPHAU                                              ;
E=J(P,P,0)                                                       ;
F=J(P,P,0)                                                       ;
*                                                                                ;
*                        SECTION 2 - PART 4                                      ;
*                                                                                ;
* THIS PART CALCULATES THE MAXIMUM LIKELIHOOD VALUES FOR PHI AND PSI             ;
* USING THE DATA AND THE CONDITIONAL VARIABLES. (M-STEP). THIS PROGRAM           ;
* WILL KEEP LOOPING TO THE LAST PART    UNTIL THE DIFFERENCES IN PHI              ;
* AND PSI, AND THE NEW ESTIMATES OF PHI AND PSI ARE LESS THAN .01.               ;
*                                                                                ;
DO I=1 TO K                                                      ;
C =(P*I)-P+1                                                     ;
D =P*I                                                           ;
G =Q(C:D,)*(W+(YM(C:D,)-U)*(YM(C:D,)-U)')*Q(C:D,)-Q(C:D,)   ;
E =E+1#/NK(I,)*G                                                 ;
F =F+G                                                           ;
END                                                              ;
FREE A                                                           ;
EE=(B-(N-K)*PSI)                                                 ;
ONE1=(PHI*((1#/K)*F)*PHI)                                        ;
PH1 =PHI+ONE1                                                    ;
TWO1=((1#/N)*(PSI*E*PSI+EE))                                     ;
PS1 =PSI+TWO1                                                    ;
PH1D=DIAG(PH1)                                                   ;
PS1D=DIAG(PS1)                                                   ;
PH1D=PH1D<>ZERO1                                                 ;
PS1D=PS1D<>ZERO1                                                 ;
PHI = PH1-DIAG(PH1)+PH1D                                         ;
PSI = PS1-DIAG(PS1)+PS1D                                         ;
FREE PS1 Q W  PH1 PH1D PS1D G E F EE                             ;
IF BUDDY GT 25 THEN GO TO FINAL;
IF MAX(ABS(ONE1)) LT 0.01 AND MAX(ABS(TWO1)) LT 0.01
   THEN GO TO FINAL                                              ;
   ELSE GO TO BUD                                                ;
FINAL:PRINT  BUDDY PHI PSI                                       ;
FREE PS1 ONE1 TWO1 U BUDDY                                       ;
*                                                                                ;
*                                                                                ;
* END OF SECTION 2                                                               ;
*                                                                                ;
*                        SECTION 3 - PART 1                                      ;
*                                                                                ;
* THIS SECTION USES THE EM ALGORITHM TO GET ESTIMATES OF THE                     ;
* RESTRICTED MODEL.  PH, OM, PS1 AND PS2 ARE ESTIMATED WITH STRUCTURE            ;
* APPLIED TO THE MODEL.  THIS FIRST PART MAKES USE OF PHI AND PSI FROM           ;
* THE LAST SECTION TO GET OPENING ESTIMATES OF PH, OM, PS1 AND PS2.              ;
*                                                                                ;
```

```
Y1=INV(L'*L)                                                    ;
Y2=INV(LA'*LA)                                                  ;
PH=Y1*L'*PHI*L*Y1                                               ;
OM=Y2*LA'*PSI*LA*Y2                                             ;
PS1=PSI-LA*OM*LA'                                               ;
PS2=PHI-L*PH*L'                                                 ;
*NOTE 'THESE ARE THE STARTING VALUES IN THIS STEP'              ;
*PRINT PH OM PS1                                                        ;
*                                                                       ;
*                       SECTION 3 - PART 2                              ;
*                                                                       ;
* THIS PART CREATES THE CONDITIONAL VALUES FOR THE MEAN AND GROUP       ;
* EFFECT FOR PH, OM, PS1 AND PS2.  THERE ARE FOUR IMPORTANT VARIABLES   ;
* CREATED HERE.   THEY ARE CREATED IN SUBROUTINES BETAB AND BETAU.      ;
* THIS IS PART OF THE INTERATIVE LOOP, THE E STEP.                      ;
*                                                                       ;
BUDDY1=0                                                                ;
BUD1: MM=(LA*OM*LA'+PS1)                                                ;
AA=(L*PH*L'+PS2)                                                        ;
BUDDY1=BUDDY1+1                                                         ;
M=INV(MM)               ; *INV OF WITHIN VARIANCES       -  P X P      ;
IF NROW(PAT) EQ 1 THEN LINK BETAB                                       ;
   ELSE LINK BETAU                                                      ;
*                       SECTION 3 - PART 3                              ;
*                                                                       ;
* THIS PART CALCULATES THE MAXIMUM LIKELIHOOD VALUES FOR PH OM PS1 PS2  ;
* USING THE DATA AND THE CONDITIONAL VARIABLES  (M-STEP). THIS PROGRAM  ;
* WILL KEEP LOOPING TO THE LAST PART UNTIL THE DIFFERENCES IN PH, OM,   ;
* PS1 AND PS2 AND THEIR NEW ESTIMATES ARE LESS THAN .01.                ;
*                                                                       ;
DO I=1 TO K             ;
C=(P*I)-P+1             ;
D=P*I                                                                   ;
LV=PH*L'*Q(C:D,)*(YM(C:D,)-U)   ;
TH=TH//LV               ;
END                     ;
FREE LV                 ;
CVE= J(P,P,0)           ;
BVE= J(P,P,0)           ;
DO I=1 TO K             ;
C=(P*I)-P+1             ;
D=P*I                   ;
CC=(S*I)-S+1            ;
DD=S*I                  ;
Z=L*TH(CC:DD,)+U        ;
AVE=NK(I,)*M*(SS(C:D,)-(YM(C:D,)*Z')-(Z*YM(C:D,)')+(Z*Z'))*M  ;
BVE=BVE+(Q(C:D,)*( (YM(C:D,)-U) * (YM(C:D,)-U)'+W)*Q(C:D,)-Q(C:D,) );
CVE=1#/NK(I,)*( Q(C:D,)*W*Q(C:D,)-Q(C:D,) ) +CVE +AVE ;
END                                                             ;
E=(N-K)*M                                                       ;
ONE=((1#/K)*(PH*L'*BVE*L*PH))                                    ;
TWO=((1#/N)*OM*LA'*(CVE-E)*LA*OM)                                ;
THREE=DIAG((1#/K)*(PS2*BVE*PS2))                                ;
```

```
FOUR=DIAG((1#/N)*(PS1*(CVE-E)*PS1))                                    ;
PH=PH+ONE                                                              ;
OM=OM+TWO                                                              ;
PS1=DIAG(PS1+FOUR)                                                     ;
PS2=DIAG(PS2+THREE)                                                    ;
PHD=DIAG(PH)<>ZERO2                                                    ;
OMD=DIAG(OM)<>ZERO3                                                    ;
PSA=DIAG(PS1)<>ZERO1                                                   ;
PSB=DIAG(PS2)<>ZERO1                                                   ;
PH=PH-DIAG(PH)+PHD                                                     ;
OM=OM-DIAG(OM)+OMD                                                     ;
PS1=PS1-DIAG(PS1)+PSA                                                  ;
PS2=PS2-DIAG(PS2)+PSB                                                  ;
FREE    Q W CVE AVE BVE TH                                             ;
IF BUDDY1 GT 250 THEN GO TO FINAL1                                     ;
IF MAX(ABS(ONE)) LT 0.01 AND MAX(ABS(TWO)) LT 0.01 AND
   MAX(ABS(THREE)) LT 0.01 AND
   MAX(ABS(FOUR)) LT 0.01 THEN GO TO FINAL1                            ;
   ELSE GO TO BUD1                                                     ;
FINAL1: PRINT BUDDY1                        PH OM PS1 PS2              ;
FREE NK RD YM SS N K R S                                          ;
IF CIRCLE LT 3 THEN GO TO BEGIN                                   ;
PRINT SEED                                                        ;
STOP                                                              ;
*                                                                 ;
*  HERE ARE THE SUBROUTINES                                       ;
*                                                                 ;
*                ALPHAU                                           ;
*                                                                 ;
ALPHAU:  TOT =J(P,P,0)                                            ;
DO I=1 TO K                                                       ;
T1 =INV((PSI*(1#/NK(I,)))+PHI)                                    ;
TOT =TOT+T1                                                       ;
Q =Q//T1                                                          ;
END                                                               ;
W =INV(TOT)                                                       ;
U =W*Q'*YM                                                        ;
FREE  T1 TOT                                                      ;
RETURN                                                            ;
*                                                                 ;
*                ALPHAB                                           ;
*                                                                 ;
ALPHAB:   T1 =INV((PSI*(1#/NK(1,)))+PHI)                          ;
Q  =GRP @ T1                                                      ;
W  =INV(T1*K)                                                     ;
U =W*Q'*YM                                                        ;
FREE  T1                                                          ;
RETURN                                                            ;
*                                                                 ;
*                BETAU                                            ;
*                                                                 ;
BETAU:  W = J(P,P,0)                                              ;
DO I=1 TO K                                                       ;
```

```
A=INV( AA + (MM*(1#/NK(I,)))  )                                    ;
Q=Q//A                    ; *MATRIX OF Q       - KP X P;
W=W+A                                                              ;
END                                                                ;
FREE A                                                             ;
W=INV(W)                  ; *COND VAR FOR U    -  P X P;
U=W*Q'*YM                 ; *COND U            -  P X P;
RETURN                                                             ;
*                                                                  ;
*                BETAB                                             ;
*                                                                  ;
BETAB:  A=INV( AA + (MM*(1#/NK(1,)))  )                            ;
Q  =GRP @ A                                                        ;
W  =INV(A*K)                                                       ;
U=W*Q'*YM                 ; *COND U            -  P X P;
RETURN                                                             ;
*                                                                  ;
*                                                                  ;
```

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, T. W.,(1984). An Introduction to Multivariate Statistical Analysis, Second Edition. John Wiley and Sons, New York.

Anderson, T. W. and Rubin, D. B.,(1956). Statistical inference in factor analysis, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. (Neyman, ed.), Vol. V, University of California, Berkeley, California.

Ahrens, H.,(1978). MINQUE and ANOVA estimator for one-way classification - a risk comparison. Biometric Journal, 19, 535-556.

Ahrens, H., Kleffe, J. and Tenzler, R.,(1981). Mean Square Error comparison for MINQUE, ANOVA and two alternative estimators under the unbalanced one-way random model. Biometric Journal, 23(4), 323-342.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N.,(1970). A maximimization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Annul of Mathematical Statistics, 41, 146-147.

Bock, R. D.,(1960). Components of variance analysis as a structural and a discriminal analysis for psychological tests. British Journal of Statistical Psychology, 13, 151-163.

Bock, R. D. and Bargmann, R. E.,(1966). Analysis of covariance structures. Psychometrika, 31, 507-534.

Brown, M. L.,(1974). Identification of the sources of significance in two-way tables. Applied Statistics, 5, 405-413.

Burstein, L., Linn, R. L., and Capell, F. J.,(1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. Journal of Educational Statistics, 3(4), 347-383.

Burt, Cyril,(1947). Factor analysis and analysis of variance. British Journal of Psychology. Statistic Section, 1, 3-26.

Chatterjee, S. K. and Das, K,(1983). Estimation of variance components for one-way classification with heteroscedastic error. Calcutta Statistical Association Bulletin, 32, 57-78.

Cronbach, L. J., with the assistance of J. E. Deken and N. Webb, (1976). Research on classrooms and schools: formulation of questions, design, and analysis. Occasional paper, Stanford Evaluation Consortium.

Dempster, A. P.,Laird, N. M.,and Rubin, D. B.,(1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B, 39(1), 1-38.

Hartley, H. O.,(1958). Maximum Likelihood estimation from incomplete data. Biometrics, 14, 174-194.

Hartley, H. O. and Hocking, R. R.,(1971). The analysis of incomplete data. Biometrics, 27, 783-808.

Healy, M. and Westmacott, M.,(1956). Missing values in experiments analysed on automatic computers. Applied Statistics, 5, 203-206.

Howe, W. G.,(1955). Some contributions to factor analysis. Report No. ORNL-1919. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Henderson, C. R.,(1953). Estimation of variance and covariance components. Biometrics, 9, 226-252.

Joreskog, K.G.(1967). Some contributions to maximum likelihood factor analysis. Psychometrika, 32, 443-482.

Joreskog, K.G.(1966). Testing a simple structure hypothesis in factor analysis. Psychometrika, 31(2), 165-178.

Joreskog, K.G.(1969). Efficient estimation in image factor analysis. Psychometrika, 34, 51-75.

Joreskog, K.G.(1970). A general method for analysis of covariance structures. Biometrika, 57, 239-251.

Joreskog, K.G.(1971). Statistical analysis of sets of congeneric tests. Psychometrica, 36(2), 109-133.

Keesling, J. W., and Wiley, D. E.,(1974). Regression models for hierarchial data. Paper presented at the Annual Meeting of the Psychometric Society, Stanford University.

Kleffe, J.,(1977). Best unbiased estimators for variance components with application to the unbalanced one-way classification. Biometrics, 32, 793-804.

La Motte, L. R.,(1973). Quadratic estimation of variance components. Biometrics, 29, 311-330.

La Motte, L. R.,(1976). Invariant quadratic estimators in the random one-way ANOVA model. Biometrics, 32, 793-804.

Lawley, D. N.,(1958). Estimation in factor analysis under various initial assumptions. British Journal of Statistical Psychology, 11, 1-12.

Lord, F. M. and Novich, M. R.,(1968). Statistical Theories of Mental Test Scores, Addison-Wesley Publishing Company Inc., Massachusetts.

Morrison, D., F.,(1967). Multivariate Statistical Methods, McGraw-Hill, New York.

Rao, C. R.,(1971). Estimation of variance and covariance components - MINQUE theory. Journal of Multivariate Analysis, 1, 257-275.

Rao, C. R.,(1972). Estimation of variance and covariance components in linear models. Journal of American Statistical Association, 67, 112-115.

Raudenbush, S. W.,(1986). Applications of a Hierchical Linear Model in Educational Research. Unpublished Doctorial dissertation, Graduate School of Education, Harvard.

Schmidt, W., C.(1969). Covariance structure analysis of the multivariate random effects model. Unpublished doctorial dissertation. University of Chicago.

Searle, S. R.,(1971). Topics in variance components estimation. Biometrics, 27, 1-76.

Spearman, C.,(1904). General intelligence, objectively determined and measured. American Journal of Psychology, 15, 201-293.

Tiao, G. C. and Tan, W. Y.,(1965). Baysean analysis of random-effect models in the analysis of variance, I. Posterior distributions of variance components. Biometrika, 52, 37-53.

Welch, B. L.,(1937). The significance of the difference between two means when the population variances are unequal. Biometrika, 29, 350-362.

Wiley, D. E.,(1967). Analysis of covariance structures. N.S.F. Research Grant Application.

Wiley, D. E., Schmidt W. H. and Bramble W. J.,(1973). Studies of a class of covariance structure models. Journal of American Statistical Association, 68, 317-323.

Williams, E. R., Radcliffe, D. and Speed, T. P.,(1981). Estimating missing values in multi stratum experiments. Applied Statistics, 30, 71-72.

Wisenbaker, J.,(1980). Structural equation model applied to hierchial data. Unpublished doctorial dissertation. Department of Education, Michigan State University.

Wu, C. F.,(1983). On the convergence properties of the EM algorithm. Annals of Statistics, 11, 95-103.