DOCUMENT RESUME

ED 364 552 TM 020 633

AUTHOR	Mok, Magdalena M. C.; McDonald, Roderick P.
TITLE	Alternative Models for Bilevel Factor Analysis.
SPONS AGENCY	Macquarie Univ., North Ryde (Australia).
PUB DATE	Apr 93
CONTRACT	MURG-20623421BAL
NOTE	26p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
AVAILABLE FROM	Magdelena Mok, School of Education, Macquarie University, New South Wales 2109, Australia.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comparative Analysis; Educational Environment; Elementary Secondary Education; *Estimation (Mathematics); *Factor Analysis; Foreign Countries; Goodness of Fit; *Models; Research Methodology; *Scaling; *Selection
IDENTIFIERS	*Bilevel Factor Analysis

ABSTRACT

The aim of this paper is to: (1) discuss the characteristics of seven bilevel factor models; and (2) arrive at a list of empirically based recommendations for selecting appropriate bilevel factor models, the assumptions underlying these decisions, and the consequences for interpretation of the Spearman case. The five models considered in the context of empirical data on school culture are: (1) no scaling, loadings constrained across levels; (2) no scaling, distinct loadings permitted; (3) scaling at both levels, constrained equal loadings; (4) scaling at both levels, constrained equal loadings, estimating school-level factor variance; and (5) scaling at both levels, distinct loadings. Analyses suggest that choice among models can depend on whether distinct factor structures across levels can be interpreted meaningfully, whether standardization of variables is necessary, and whether the researcher wants to estimate second level factor variance. If distinct factor structures are interpretable, then isomorphic models 2 and 5 would provide the best fit, and would, in most cases, provide no better fit than the more parsimonious model 4. Parsimony and ease of interpretation favor model 4. Model 1 can provide a direct estimate of school effect not available from other models. Each of the models, except for model 3, provides useful and complementary information. Three figures and two tables illustrate the discussion. (Contains 25 references.) (SLD)

Alternative Models for Bilevel Factor Analysis

Magdalena M. C. Mok
Macquarie University

&

Roderick P. McDonald
University of Illinios

Paper presented at the 1993 Annual Conference
of the American Education Research Association,
12 - 16 April, Atlanta, Georgia, U.S.A.

Acknowledgments

The research of the first author was partially supported by the Macquarie University MURG Grant No. 20623421BAL. The authors would like to thank our research assistant Miss Alison Warton for her efficient work, Miss Jessica Hammond for her excellent typing, and Br Marcellin Flynn for allowing us to use the set of data for this study.

Requests for reprints should be directed to:
Dr. Magdelena Mok, School of Education, Macquarie University, NSW 2109, Australia.

Alternative Models for Bilevel Factor Analysis

Magdalena M.C. Mok & Roderick P. McDonald

Introduction

Educational research often involves collecting data that have a nested sampling structure. For example students are randomly selected from a random selection of schools. As well, educational research often addresses questions that are at the institutional level while data are collected only at the student level. For example, studies on school culture were often based on questionnaire data collected from students. It has been established by recent developments in research methods that hierarchical structural models are to be preferred over conventional data, or in designs involving research questions at the institutional levels. (See for example, Aitkin & Longford, 1986; Bock, 1989; Bryk & Raudenbush, 1992; Goldstein, 1987; Mason, Wong & Entwisle, 1983; Raudenbush & Bryk, 1985.)

In the area of structural equation models, discussions emanating from Goldstein & McDonald's (1988) general model for bilevel structures include developments on the statistical theory (Lee, 1990) and validations of the model on empirical data sets (e.g. Muthen, 1989; 1991).

Nevertheless, the literature is still unclear as to the selection of appropriate multilevel factor models from those resulting from (i) decisions on scaling at the various levels, and (ii) the possibility of equating or not equating factor loadings across levels. It is the aim of this paper to (a) discuss the characteristics of seven bilevel factor models, and (b) arrive at a list of empirically based recommendations for selecting appropriate bilevel factor models, the assumptions underlying these decisions, and the consequences for interpretation for the Spearman Case.

Theoretical Framework

The models discussed in this paper are special cases of a bilevel model given by McDonald and Goldstein (1989), in which measures are obtained at the student level only. A q x 1 data vector of measures on student i randomly sampled from school j, also randomly sampled, $\underline{y}_{ij}$ ($i=1, \ldots, n_j$, $j=1, \ldots, m$) is given a variance component decomposition

$$\underline{y}_{ij} = \underline{\mu} + \underline{y}_{(B)j} + \underline{y}_{(W)ij}$$

where μ is the grand mean, $\underline{y}_{(B)j}$ is the between - school component and $\underline{y}_{(W)ij}$ is the within - school component. The bilevel factor model is, using results from McDonald and Goldstein (1989),

$$\underline{y}_{ij} = \underline{\lambda}_j \underline{z}_{ij} + \underline{u}_{ij}$$

or

$$\underline{y} = \Lambda \underline{z} + \underline{u}$$

where $\underline{z}$ is the factor score, $\underline{u}$ is the unique component and Λ is the factor loading matrix, and is given by

$$\Lambda = \bigoplus_{j=1}^{m} (I_{N_j} \otimes \underline{\lambda}_j)$$

The covariance matrix, $\sum_T$ of $\underline{y}$ is decomposed into between - and within - school components $\sum_B$ and $\sum_W$ respectively, as

$$\sum_T = \sum_B + \sum_W$$

such that

$$\sum_B = \Lambda_B \Theta_B \Lambda'_B + \Psi_B$$

$$\sum_W = \Lambda_W \Theta_W \Lambda'_W + \Psi_W$$

where the Λ's are the factor loading matrices, Θ's are factor covariance matrices and Ψ's are diagonal residual covariance matrices, with the subscripts indicating the corresponding levels. By equating or nor equating Λ_B and Λ_W, and by choosing various scaling factors on Θ_B and Θ_W, different structural models, some of which may be shown to be reparameterizations of others, can be fitted. These are discussed in more detail below, in the context of a Spearman Case.

I. Model A

In this model, the original matrix of measurements is used and the factor loadings are constrained equal across the two levels. By decomposing the factor score and the unique factor of the data vector into a between - school and a within - school component, this model can be represented by:

$$\underline{y}_{ij} = \underline{\lambda}(\eta_j + \zeta_{ij}) + (\underline{\nu}_j + \underline{\varepsilon}_{ij})$$
$$= (\underline{\lambda}\eta_j + \underline{\nu}_j) + (\underline{\lambda}\zeta_{ij} + \underline{\varepsilon}_{ij})$$
$$= \underline{y}_{(B)j} + \underline{y}_{(W)ij}$$

where η and ζ are the school-level and student-level factor scores and ν and ε are the school- and student-level unique factor respectively.

The corresponding covariance matrix $\sum_T$ is then decomposed into the between - and within - school components:

$$\sum_T = \sum_B + \sum_W$$

where $$\sum_B = \underline{\lambda}\sigma_B^2\underline{\lambda}' + \Psi_B$$

$$\sum_W = \underline{\lambda}\underline{\lambda}' + \Psi_W$$

and σ_B^2 is the between - school factor variance.

This model has the simple interpretation that there is between - school variability and within - school variability in the common factor and the unique factors, yielding the corresponding between / within variability in the item scores. In this model, the between - school factor variance is directly estimated. The ratio σ_B^2 / (σ_B^2 + 1) provides a natural estimate of school effects. The limitation of this model is the constraint on the equality of factor loadings across levels, with the implication that the same factor underlies the behaviour of items at the school- and the student - levels. These points can be elaborated using a six - item scale designed to measure the quality of the formal curriculum (QUAL). The items in this scale are

(1) The curriculum of this school meets my present needs.
(2) This school offers a good range of subjects in Years 11 and 12.
(3) The subjects offered at this school develop the capacity for independent and critical thinking.
(4) The subjects taught in this school offer useful knowledge or skills.

(5) The subjects taught in this school are relevant to real life and to students' needs.
(6) The subjects taught here prepare students adequately for future employment.

The semantics of these items suggest that item 2 refers more to the curriculum at the institutional level while item 1 is more related to the students' personal opinion of the school curriculum. The rest of the items in the scale are expected to be impacted upon by a mixture of school-level and student-level variations. As such, the constrained-equal loading model might not be the best to use. Schematically, model A for the QUALity of the formal Curriculum Scale can be represented by Figure 1 below.

Place Figure 1 about here

Relaxing the equal loading constraint leads naturally to the next model below.

II. Model B

This model relaxes the assumption on equal loadings across levels of model A. The previous metric is still used in the analysis. This model, using the same notation as before, is:

$$
\begin{aligned}
\underline{y}_{ij} &= (\underline{\lambda}_B \eta_j + \underline{\lambda}_W \zeta_{ij}) + (\underline{\nu}_j + \underline{\varepsilon}_{ij}) \\
&= (\underline{\lambda}_B \eta_j + \underline{\nu}_j) + (\underline{\lambda}_W \zeta_{ij} + \underline{\varepsilon}_{ij}) \\
&= \underline{y}_{(B)j} + \underline{y}_{(W)ij}
\end{aligned}
$$

and $\qquad \sum_T = \sum_B + \sum_W$

where

$$
\begin{aligned}
\sum_B &= \underline{\lambda}_B \underline{\lambda}'_B + \psi_B \\
\sum_W &= \underline{\lambda}_W \underline{\lambda}'_W + \psi_W
\end{aligned}
$$

The price for allowing factor loadings to freely vary across the two levels is that, in the usual sense in single-level factor analysis, factor variances at both levels need to be fixed at unity for identifiability. The consequence is that there is no direct way of estimating school effects parallel to that in Model A. It is possible, on the other hand, to compute the ratio of

$$\underline{\lambda}'_{*}\underline{\lambda}_{*} / (\underline{\lambda}'_{*}\underline{\lambda}_{*} + \underline{\lambda}'_{**}\underline{\lambda}_{**})$$

as an estimate of the relative amount of information contributed by the school-level variations. Nevertheless, the more severe consequence of this distinct loading model is the implication that, in cases where the loadings are significantly different across levels, the factors at the school- and student-level are different. In other words, the latent attribute of the school as registered through students' perceptions may have a different interpretation to the latent attribute of the student registered through the same group of students with the same set of measuring instruments. Just for the sake of argument, suppose the school-level loadings in the Quality of formal school curriculum were "defined by" items 2, 3, 4 such that the school-level factor was one of quality in terms of subject range and levels of academic content, while student-level loadings suggested a factor defined by how well student needs were met by the curriculum. The factors measured by these items would then be completely different traits at the school- and student - levels, It would therefore be misleading to label both with the same factor name and any computation of proportion of variance explained by school versus by student would be erroneous. Of course the caveat extends itself naturally to the possibility of different factor structures in terms of number of factors across the two levels. Model B is represented schematically by Fig. 2.

Place Figure 2 about here

The above arguments are all very well except for a problem with the scaling of the units of measurement. In the same sense in which we are not usually willing to compare unstandardized regression weights, if the spread of the indicators of a factor are different, it is not possible to compare relative contributions of these indicators in defining the factor by comparing the factor loadings; without appropriate scaling, the loadings represent the weights of these indicators measured in their natural units. This necessitates the development of the next set of models.

III. Model C

The variables are scaled to yield standardized factor loadings at both levels first for this model. Like Model A, the loadings in Model C are constrained equal across levels. In single-level factor analysis, scale invariant models are obtained by pre- and post- multiplication of a structured correlation matrix by a non - singular diagonal scaling matrix, Δ, which is to be estimated (Krane & McDonald, 1978. See also Brown, 1982; McDonald, Parker & Ishizuka, in press).

The mathematical expressions defining Model C are:

$$\underline{y}_{(B)j} = \Delta_B \underline{g}_{(B)j} = \Delta_B(\underline{\alpha}\eta_j + \underline{\nu}_j)$$
$$\underline{y}_{(W)ij} = \Delta_W \underline{g}_{(W)ij} = \Delta_W(\underline{\alpha}\zeta_{ij} + \underline{\varepsilon}_{ij})$$

where Δ_B and Δ_W are the scaling matrices at the between - and within - levels respectively and are to be estimated. Extending the arguments for the single-level case given in McDonald, Parker, and Ishizuka (in press) to the bilevel situation, $\sum_{g(B)}$ and $\sum_{g(W)}$ are both constrained such that Diag $\{P_B\}$=I and Diag $\{P_W\}$=I , in

$$\sum_B = \Delta_B \mathcal{P}_B \Delta_B \quad \text{and} \quad \sum_W = \Delta_W \mathcal{P}_W \Delta_W$$

by

$$\mathcal{P}_B = \underline{\alpha}\underline{\alpha}' + \mathit{Diag}\{I - \underline{\alpha}\underline{\alpha}'\}$$
$$\mathcal{P}_W = \underline{\alpha}\underline{\alpha}' + \mathit{Diag}\{I - \underline{\alpha}\underline{\alpha}'\}$$

yielding standardized factor loadings $\underline{\alpha}$. This model is to be preferred over Model A whenever the variance of the indicators are very different. However because the equated loadings in this Model are derived from standardised scores, comparisons can only be made in the relative sense. The constraint on holding the common factor variances at unity at both levels implies that the general size of school effects cannot be directly defined and estimated.

IV. Model D

This model is an attempt to relax the restrictions on Model C such that school-level factor variance can be estimated with scaling for standardized parameters at both levels. Model D is thus a model involving scaling at both levels and constraining loadings across levels.

This model is represented by:

$$\underline{y}_{(B)j} = \Delta_B \underline{g}_{(B)j} = \Delta_B\ (\underline{\alpha}\eta_j + \underline{\nu}_j)$$
$$\underline{y}_{(W)ij} = \Delta_W \underline{g}_{(W)ij} = \Delta_W\ (\underline{\alpha}\zeta_{ij} + \underline{\varepsilon}_{ij})$$

and

$$\Sigma_B = \Delta_B(\underline{\alpha}\sigma_B^2\underline{\alpha}' + C_B)\Delta_B$$
$$\Sigma_W = \Delta_W(\underline{\alpha}\underline{\alpha}' + C_W)\Delta_B$$

where C_B and C_W are diagonal matrices of unique variances satisfying the scaling constraints. Model D offers a direct estimate of school-level variance, viz, σ_B^2 and the corresponding school effect estimate of σ_B^2 / (σ_B^2 + 1). Estimates of school-level loadings tend to be more stable (smaller SE) compared to the distinct loadings counter parts because estimation is based on the pooled larger sample size at the student-level. In research situations where distinct loadings are expected across levels, the next model may be considered.

V. Model E

Model E is an attempt to relax the constraints on equal loadings across levels while still standardising the variables at both levels. Mathematically, this model is represented by:

$$\underline{y}_{(B)j} = \Delta_B \underline{g}_{(B)j} = \Delta_B\ (\underline{\alpha}_B \eta_j + \underline{\nu}_j)$$
$$\underline{y}_{(W)ij} = \Delta_W \underline{g}_{(W)ij} = \Delta_W\ (\underline{\alpha}_W \zeta_{ij} + \underline{\varepsilon}_{ij})$$

and

$$\Sigma_B = \Delta_B(\underline{\alpha}_B\underline{\alpha}'_B + C_B)\Delta_B$$
$$\Sigma_W = \Delta_W(\underline{\alpha}_W\underline{\alpha}'_W + C_W)\Delta_W$$

This model can be shown to be a reparameterization of Model B by writing

$$\underline{\lambda}_B = \Delta_B\underline{\alpha}_B$$

and

$$\underline{\lambda}_W = \Delta_W\underline{\alpha}_W$$

This model then is expected to give the same fit to the data as Model B and all the comments regarding Model B apply here. The additional advantage of Model E over Model B is that by means of the standardisation procedure, the loadings can be interpreted by the usual rules for standardized parameters.

VI. Other Bilevel Factor Models

Mathematically it is possible to scale for standardized loadings only at the student-level while allowing the school-level to remain at the original units of measurement if the loadings are distinct across levels:

$$\underline{y}_{(B)j} = \underline{\lambda}_B \eta_j + \nu_j$$

$$\underline{y}_{(W)ij} = \Delta_W (\underline{\alpha}_W \zeta_{ij} + \underline{\varepsilon}_{ij})$$

with

$$\Sigma_B = \underline{\lambda}_B \underline{\lambda}'_B + \Psi_B$$

and

$$\Sigma_W = \Delta_W (\underline{\alpha}_W \underline{\alpha}'_W + C_W) \Delta_W$$

Indeed by writing

$$\underline{\lambda}_B = \Delta_B \underline{\alpha}_B$$

this can be shown to be a reparameterization of Model E. Nevertheless, there does not seem to be an empirical situation which motivates this type of "mid - way" scaling. So this model will be treated as one with theoretical interest only but with very little use in terms of applications.

On the other hand, it is also possible to fit a mathematical counter part for the "Equal-Loadings-Scaling-at-Student-Level-only" model. Again, this model does not seem well motivated.

So far decompositions in all the models discussed are in terms of common factor scores and unique factor scores, with all parameters fixed. It is perfectly legitimate on the other hand to decompose the factor loadings into between - and within - components in a manner analogous to the random coefficients models discussed by de Leeuw and Kreft (1986). That is,

$$\underline{y}_{ij} = (\underline{\alpha}_0 + \underline{\beta}_j) z_{ij} + \underline{u}_{ij}$$

$$= (\underline{\alpha}_0 + \underline{\beta}_j) z_{ij} + (\underline{v}_j + \underline{\varepsilon}_{ij})$$

For the Quality of Formal Curriculum scale, this model can be represented by Figure 3.

Place Figure 3 about here

In other words, the random coefficient model implies that the loadings of the manifest variables on the common factor are being randomly "modified" by school memberships. In this model, students are assumed to have a perception of the quality of school curriculum and that this perception is being coloured by the particular school the student happen to belong to. School effect is modelled through its impact on the relationship between the latent and manifest variables. The software to be reported in this paper was not written to handle a random coefficient model and so this model will not be further discussed. Nevertheless, it is possible to give an estimate of the profile of the slopes simply by repeatedly running a single level factor analysis on each school sample if needed.

Source of Data

Illustrative data consist of responses to a questionnaire from 5,932 Year 12 students. The students were selected from 50 Catholic schools from the Dioceses of NSW, Australia, by means of stratified random sampling (Flynn, 1992). The questionnaire contains 46 Likert - type items pertaining to 6 scales which were constructed to measure school culture. The scales are: Quality of Formal School Curriculum (QUAL, 6 items), Out - of - school curriculum (OUTS, 4 items), Relationships with Teachers (ATEC, 12 items), Student Morale (STDM, 10 items), Attitudes to the Principal (PRIN, 5 items), and the Attitudes to Discipline (ATTD, 4 items). The rationale behind these scales is discussed elsewhere (Flynn, 1992; Mok, 1992) and will not be repeated here.

Methods

The strategies (with modification) for multilevel covariance structure analysis recommended by Muthen (1991) were adopted. The steps were:

Step 1. Conventional Factor Analysis of S_T

An Maximum Likelihood confirmatory factor analysis of each of the six scales was conducted by means by the COSAN computer package (Fraser, 1987). Items which have a loading less than 0.33 were removed if necessary. This step was conducted with the understanding that the analysis was incorrect in the sense that the nested nature of the data had been ignored (Muthen, 1991)

Step 2. Estimation of Between Variation

The intra - class correlation of each item in each scale was computed by means of fitting a variance components model to the data. The intra - class correlation gives the proportion of variance accounted for by the school. If it were small then multilevel analysis might not be warranted (Muthen, 1991). This step therefore was aimed to determine the legitimacy of the bilevel factor analysis to follow. The variance component model for each item was fitted by the ML3 computer software (Rasbash, Prosser, & Goldstein, 1990), which adopted an Iterative Generalised Least Squares algorithm in minimising the log - likelihood function (Goldstein, 1989).

Step 3. Bilevel Factor Analysis

The analyses reported here were performed using McDonald's BIRAM (Bilevel Reticular Action Model) software program (McDonald, Lam, Middlehurst, & Parker, in preparation). This program is written to implement the Reticular Action Model (McArdle & McDonald, 1984) for bilevel path analysis. Maximum Likelihood estimates of parameters are obtained by a quasi - Newton iterative procedure. Goodness of fit is obtained by comparing the negative log - likelihood ratios (chi - squared values) of the fitted model to a saturated model. The program provides the option for the user to set up the model for the run, or to use some existing models from previous runs. A control file was set up for each of the models discussed (Model A to E) for each of the six school culture scales. This was done by (i) selecting either scaling for standardized parameters (at both levels) or no scaling and (ii) specifying either estimating or not estimating school-level factor variances.

For each scale and within each model, factor loadings across school- and student - levels were compared. Factor loadings between models for each level were compared and interpreted. Goodness of fit (Chi-squared value) across the models on each of the scales was also compared and contrasted as well. As BIRAM is a program still under development, it is anticipated that difficulties associated with the convergence of the minimisation procedure as a consequence of the combination of small intraclass correlation and unbalanced sample sizes would occur for some of the scales. These difficulties were also summarised here for further understanding of the selection of bilevel factor models. Altogether 5 x 6 = 30 BIRAM analyses were performed.

The BIRAM program also has a provision for running a pseudo balanced analysis whereby all second level units (schools) were assumed to take an average sample size of first level units (students), or an unbalanced analysis whereby the actual sample size is used in the estimation, taking the results from the pseudo balanced run as initial values for the minimisation procedure. Where-ever available Chi-squared values from both pseudo balanced and unbalanced runs were reported here. The aim was to compare the goodness of fit between pseudo balanced and unbalanced runs for a badly unbalanced situation where school sizes ranged from 55 to 348.

Results

Step 1. Results of the Conventional Factor Analysis Of S_T

Conventional Factor Analysis using the COSAN software package (Fraser, 1987) on each of the six scales identified items not contributing to the definition of the factor. These items were inspected and subsequently removed. The remaining items were reanalysed using COSAN and the results are summarised in Table 1.

Place Table 1 about here

The results from this step indicated reasonable fit for all of the scales with chi-squared values ranging from 2770.3 (df = 54) to 24.13 (df = 2) with the corresponding McDonald's Goodness of fit Indices ranging from 0.80 to 0.998 (McDonald, 1989). The Root Mean Square residual covariances ranged from 0.01 to 0.05 . The Cronbach's Alpha as a measure of internal consistency for these scale ranged from 0.67 to 0.91.

Step 2. Results on Estimation of Between Variation

Intraclass correlations of the 41 items identified from the previous step ranged from 0.08 to 0.18 (Table 1). These figures concurred with the other studies reported on school effectiveness (Reynolds, 1992). Muthen (1991) has suggested that if intraclass correlations are too small, multilevel analysis is not warranted. Intra - class correlations in studies reported by Muthen (1991) are a lot higher . It would be interesting to compare the effects of the sizes of the intraclass correlation on the different models. This is done in a simulation study by McDonald and Mok (1993).

Step 3. Results of Bilevel Factor Analysis

The hostile characteristic of this data set with a combination of small intra - class correlations, extremely unbalanced school sizes, and a large sample size at the student-level puts the BIRAM program to a severe test. The dimensions of intermediate computational stages exceeded the current limits of the BIRAM program for the two longer (with more than 10 items) scales (Relationships with teachers, 12 items, and Student Morale, 10 items). Even for the remaining four shorter scales (with less than 10 items), some runs failed to give standard error estimates. The results of 20 BIRAM runs for these four scales are summarised in Table 2.

Place Table 2 about here

Table 2 contains the factor loadings at the school- and student - levels and factor variance estimates where appropriate for each scale. Referring to the factor loadings for the Quality of Formal Curriculum Scale (QUAL), which has a low intra class correlation of 0.08, an item-to-item comparison between the values of the student-level loadings and the single-level loadings showed very little difference between them. This was especially true for models involving scaling. On the other hand, for the Out - of - School Curriculum (OUTS), Attitudes toward the Principal (PRIN), and the Attitudes toward Discipline (ATTD) scales, where the intra-class correlations were higher, more discrepancies between student-level loadings from the various bilevel factor models and those obtained from single-level factor analysis were observed. This was especially the case for models without scaling. These observations implied that if the intra class correlation is small, and if the research question involves a concept at the student-level only (e.g. quality of school life from the students' experience) then the conventional single-level factor analysis might be considered sufficient. On the other hand, if the intra - class correlation is high, then it might be worthwhile to proceed to the bilevel factor models which provide better estimates over the single-level factor analysis.

Comparisons of school-level loadings with student-level loadings were, strictly speaking, only legitimate on Model E which involved scaling at both levels and which allowed distinct loadings across levels. For all the scales that had successful runs on BIRAM, school-level loadings were higher than student-level loadings after scaling at both levels. The proportion of information at the school-level was estimated to be 65% for the Quality of Formal Curriculum Scale (QUAL), 62% for the Out-of-School Curriculum Scale (OUTS), 61% for the Attitudes toward the Principal Scale (PRIN), and 70% for the Attitudes towards Discipline Scale (ATTD) (Table 2). It is unfortunate that both the Out-of-School Curriculum and the Attitudes towards Discipline scales, each with an intra class correlation of 0.16, involved Heywood cases (Heywood, 1931) on Model E. Muthen (1991) has warned that Heywood cases were as common in Bilevel factor analysis as in single level factor analysis.

Referring to Model E again, school-level factor loadings for all the items in the Quality of Formal Curriculum Scale (QUAL) were in the order of 0.9 while loadings at the student-level were on the order of 0.6 to 0.7. However, the within item ratio of the school - to student - levels loadings were not uniform across all items, some items had relatively higher loadings at the school-level than others had. The same remark can be made for the Attitudes towards the Principal Scale (PRIN). School-level loadings for items 2, 3, 5 were relatively higher than those at the student-level, whereas school- and student-level loadings for items 1 and 4 were similar. These made good conceptual sense on inspection of the items (see Table 1). Items 2, 3, 5 involved the leadership of the principal while items 1 and 4 were respectively involved with how approachable the principal was perceived and the importance placed by the principal on the religious nature of the school.

The size of factor loadings from the unscaled models (Models A and B) might just be a reflection of the items' variabilities. Their interpretations can be as misleading as comparing raw regression coefficients. However, Model A (unconstrained and with no scaling) provided direct estimates of school-level factor variances, which ranged from 7% to 23% (Table 2). The size of the school-level factor variance tends to associate with the size of intra class correlation, but there does not seem to be a one-to-one relationship between them.

The Chi-squared goodness of fit values of all models on all scales for both the pseudo balanced and the unbalanced runs are also given in Table 2. It is possible to make pairwise comparisons between models on goodness of fit for each individual scale using information presented in Table 3. For example, for the Quality of Formal Curriculum (QUAL), bilevel factor model A gains 14 degrees of freedom at a cost of losing 2.01 points on Chi - squared value of goodness of fit over the single-level analysis. The loss was not statistically significant at 5% level, and because in addition the research question involved one at the school-level, the move from the single-level factor analysis to Model A of the bilevel factor analysis was justified. The exercise of such pairwise comparisons across all models and all scales suggests that at least for the scales reported here, (i) if the loadings were unconstrained across levels, then scaling and non-scaling give the same goodness of fit: the models (B and E) are reparameterizations of one another; (ii) the unconstrained models (B and E) are not significantly better than constrained Model D which involves scaling at both levels and which requests school-level factor variance to be estimated. So the more parsimonous Model D is to be preferred over the other models; (iii) each of Models B, E, and D is significantly better than the constrained Model A which involves no scaling; (iv) each of these four models is better than Model C which is in turn not significantly different from the single-level factor analysis in terms of goodness of fit All bilevel factor models except Model C yield better fit to the data than the conventional factor model.

The Chi - squared values for the pseudo - balanced and the unbalanced runs were next compared. All the Chi - squared values from the pseudo - balanced runs were larger than the unbalanced runs indicating a better fit by the latter. However, only 3 of the 20 differences were larger than 7.88. The values of the loadings show that even with an unbalanced data set which had sample size ranging from 55 to 348, pseudobalanced estimates approximate the unbalanced satisfactorily.

Conclusions

The questions facing the researcher in decisions regarding bilevel factor analysis models are :

(a) Whether bilevel factor analysis is required?
(b) If a bilevel factor analysis is to be performed what are the available models?
(c) How do we choose among these models?

Various authors have written on the first question. The general advice hinges on two points: (i) whether the research question is conceptualised at the student-level (eg quality of school life as experienced by the student), or at the school-level (eg culture of the school as an institution), and (ii) whether there is enough between - variation to warrant multilevel analysis.

Once a decision is made to conduct a multilevel analysis, it is necessary for the researcher to consider the class of alternative models.

This paper discusses some of the available alternative bilevel factor models. Eight bilevel factor models were proposed, five of them were discussed in greater detail just for the Spearman case. Selections of these models are discussed in the context of a set of empirical data on school culture. The five models are

(A) no scaling, loadings constrained equal across levels,
(B) no scaling, distinct loadings permitted,
(C) scaling at both levels, constrained equal loadings,
(D) scaling at both levels, constrained equal loadings, estimating school-level factor variance,
(E) scaling at both levels, distinct loadings.

Analyses in this study suggest that choices amongst the models depend on:
(i) whether distinct factor structures across levels can be interpreted meaningfully,
(ii) whether standardisation of variables is necessary, and
(iii) whether the researcher wants to estimate second level factor variance.

It appears from the analyses performed on the current set of empirical data, that if distinct factor structures are interpretable, then the isomorphic Models B and E would provide best fit to the data, and in most cases, these distinct models would yield no better fit than the more parsimonious model D. Because of parsimony and ease of interpretation, Model D is preferred. On the other hand, while the factor loadings from the non scaled models are not easily interpreted, Model A with constrained equal loadings provides a direct estimate of school effect that is not available from any other models. In this sense, except the very restrictive Model C, each of the other Bilevel factor models provides useful and in fact complementary pieces of information.

Results of the runs also indicate that the pseudo balanced estimates are quite robust with regard to the extent of unbalance of the sample structure. How much of this robustness depends on the total sample size is discussed elsewhere (see McDonald and Mok, 1993 for a simulated study). Likewise, the robustness of these models with respect to the sizes of the intraclass correlations cannot be fully assessed without conducting a simulation study.

So far discussions have been restricted to a one - factor Spearman case. Future research could be extended in the direction of different factor structures across levels. The mathematical models discussed in this paper placed no restrictions on the number of factors and the arguments can be extended easily to more complicated structures.

Acknowledgments

The research of the first author was partially supported by the Macquarie University MURG Grant No. 20623421BAL. The authors would like to thank our research assistant Miss Alison Warton for her efficient work, Miss Jessica Hammond for her excellent typing, and Br Marcellin Flynn for allowing us to use the set of data for this study.

References

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. Journal of the Royal Statistical Society, Series A, 149 (1), 1-43.

Bock, D. (1989). Multilevel Analysis of Educational Data. Academic Press.

Browne, M.W. (1982). "Covariance structures." In Hawkins, D.M. (ed.) Topics in Applied Multivariate Analysis. Cambridge University Press: Cambridge.

Bryk, A.S. & Raudenbush, S.W. (1992). Hierarchical Linear Models. Sage Publications.

Flynn, M. (1993). The Culture of Catholic Schools. N.S.W.: St. Paul Publications.

Fraser, C. (1987). COSAN User's Guide.

Goldstein, H. (1987). Multilevel Models in Educational and Social Research. London: Griffin; New York: Oxford University Press.

Goldstein, H. (1989). "Restricted (unbiased) iterative generalised least squares estimation." Biometrika,76, 622-623.

Goldstein, H. & McDonald, R.P.(1988). "A general model for the analysis of multilevel data." Psychometrika, 53, 455-467.

Heywood, H.B. (1931). On finite sequences of real numbers. Proceedings of the Royal Society, Series A,134, 486-501.

Krane, W.R. & McDonald, R.P. (1978). "Scale invariance and the factor analysis of correlationmatrices." Br. J. Math. Statist. Psychol, 31, 218-228.

Lee, S.Y. (1990). "Multilevel analysis of structural equationmodels." Biometrika, 77 (4), 763-72.

Mason, W.M., Wong, G.M. & Entwistle, B. (1983). "Contextual analysis through the multilevel linear model." In S. Leinhardt (Ed.), Sociological Methodology. 72-103.

McArdle, J.J. & McDonald, R.P. (1984). "Some algebraic properties of the Reticular Action Model for moment structures". British Journal of Mathematical and Statistical Psychology, 37, 234-251.

McDonald, R.P. (1989). "An Index of Goodness-of-Fit based on noncentrality". Journal of Classification, 6, 97-103.

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

McDonald, R.P., MiddleHurst, J., Lam, P., & Parker, P. (in preparation). User's Guide for the RAM package.

McDonald, R.P, & Mok, M. (1993). "The behaviour of bilevel factor models as a function of properties of the data.". Paper presented at the Annual Conference of the American Educational Research Association, 12-16 April, Atlanta, Georgia, U.S.A.

McDonald, R.P., Parker, P., & Ishizuka, T. (in preparation). "A scale-invariant treatment for recursive path models".

Mok, M. (1992). "A multilevel analysis of school culture in New South Wales". Paper presented at the AARE/NZARE Joint Conference, 22-26 November, Deakin University, Geelong, Victoria, Australia.

Muthen, B.O. (1989). "Latent variable modeling in heterogenous populations." Psychometrika, 54,557-585.

Muthen, B.O. (1991). "Multilevel Covariance Structure Analysis." in Hox, J. & Kreft I. (eds.). Multilevel Modeling. A special issue of Sociological Methods & Research.

Rasbash, J., Prosser, R., & Goldstein, H. (1990). User's Guide for the ML3 Software for Three-level Analysis. Institute of Education, University of London.

Raudenbush, S.W., & Bryk, A.S. (1986). "A hierarchical model for studying school effects." Sociology of Education, 59, 1-17.

Reynolds, D., & Cuttance, P. (1992) (Editors). School Effectiveness. London: Cassell.

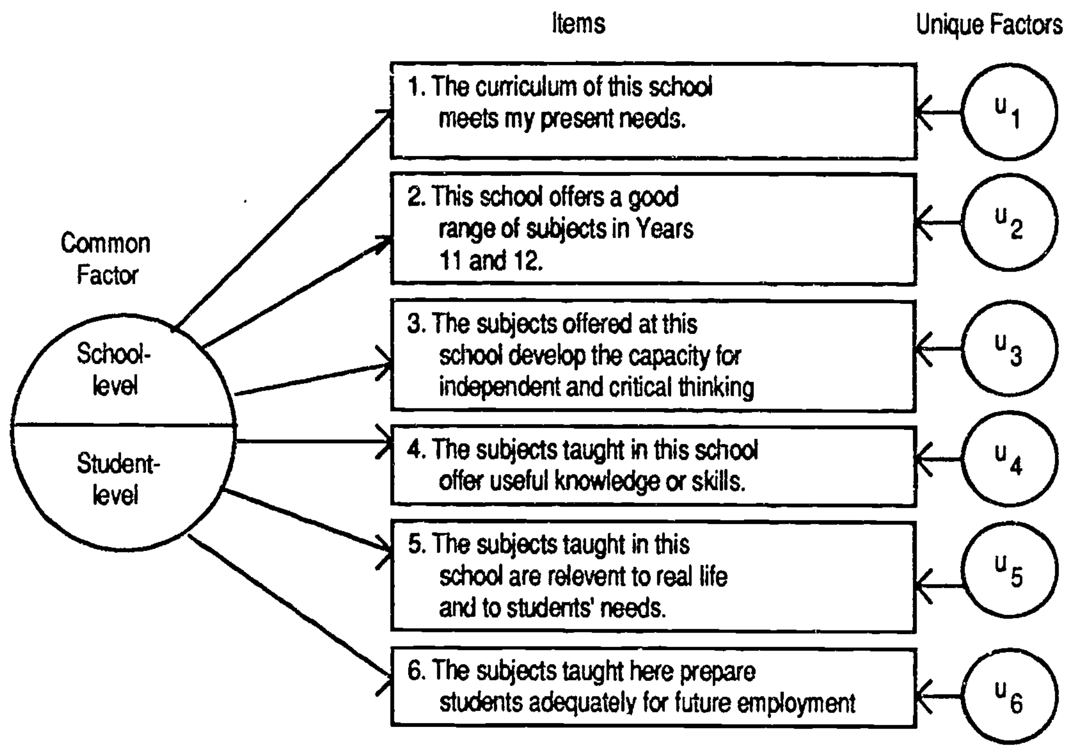

Figure 1. Model A: Equal loadings across levels; no scaling is used.

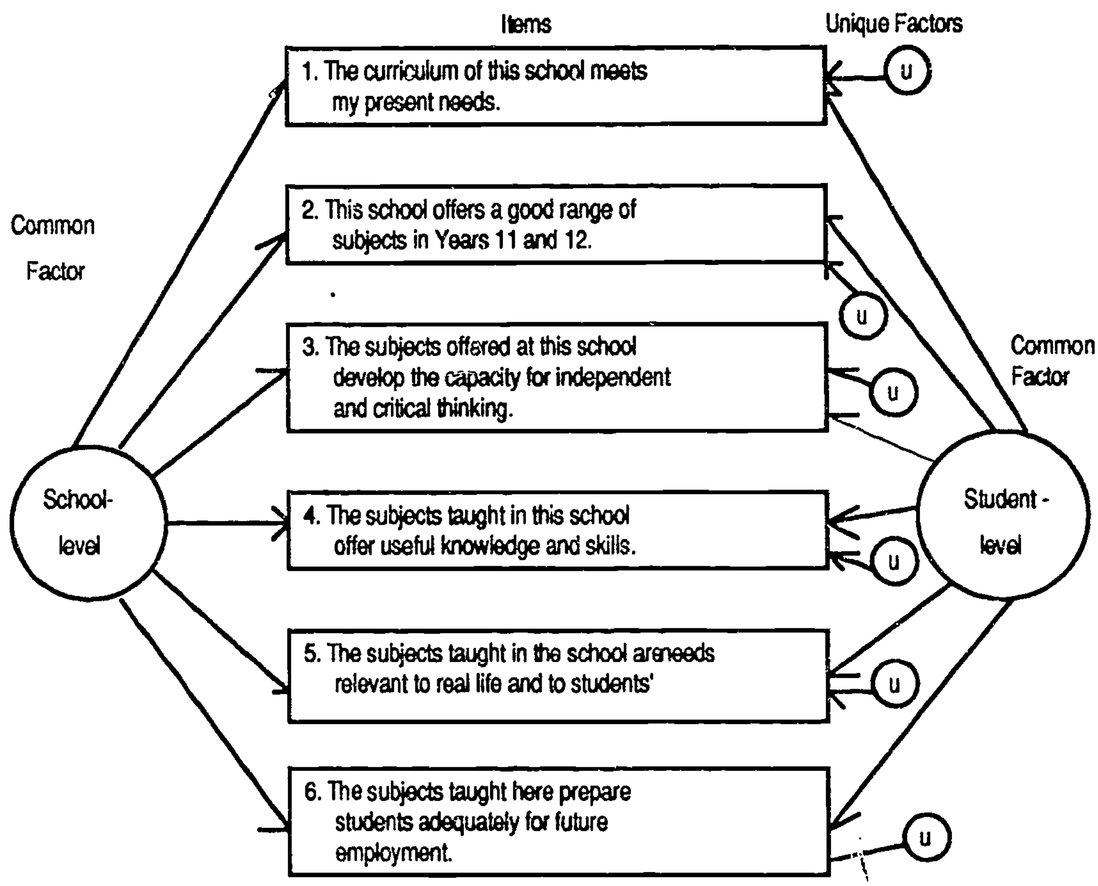

Figure 2. Model B: Unconstrained loadings across levels; no scaling is used.

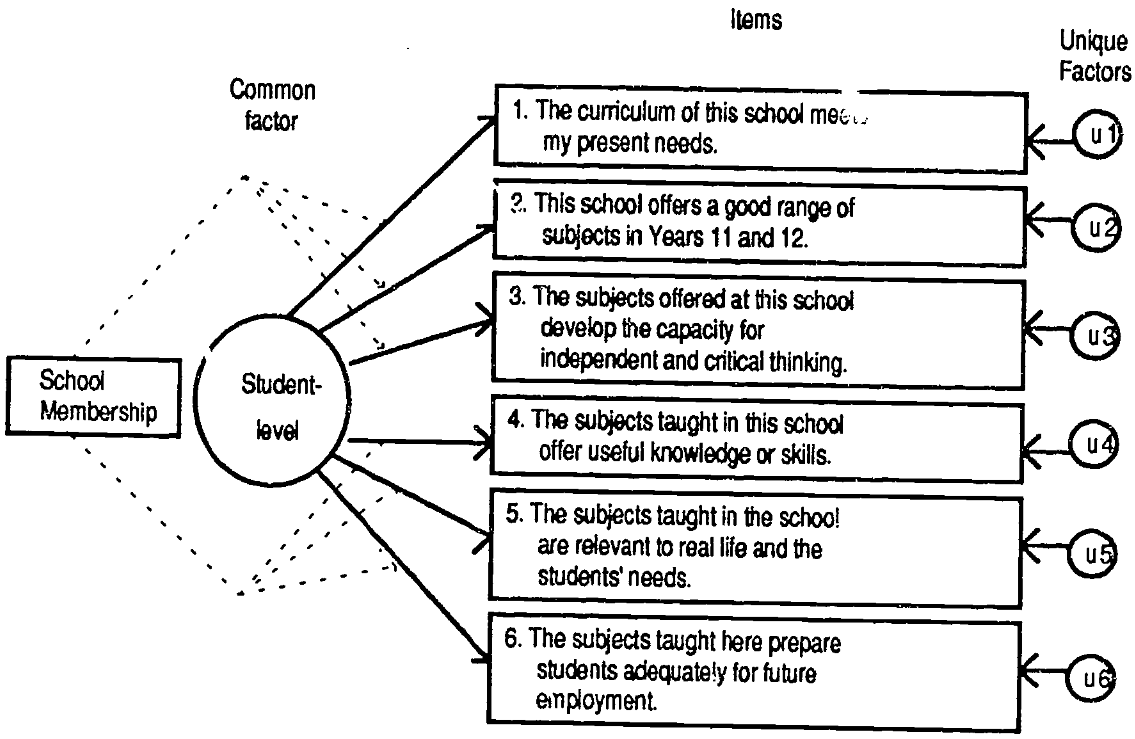

Figure 3. Random Coefficient Model

Table 1. Loadings from Confirmatory Single-level Factor Analysis and Intraclass correlations(1)

No	Item Stem	Loading(2)	Intra corr.
	QUAL - Quality of the Formal Curriculum (Chi-sq = 755.77, df = 9, McDonald's Index = 0.94)		
41	The curriculum of this school meets my present needs.	.64 (.01)	.04
45	This school offers a good range of subjects in Years 11 and 12.	.64 (.01)	.16
46	The subjects offered at this school develop the capacity for independent and critical thinking.	.72 (.01)	.06
47	The subjects taught in this school offer useful knowledge or skills.	.74 (.01)	.03
49	The subjects taught in the school are relevant to real life and to students' needs	.66 (.01)	.02
50	The subjects taught here prepare students adequately for future employment	.66 (.01)	.04
	OUTS - Out-of-School Curriculum (Chi-sq = 68.08, df = 2, McDonald's Index = 0.99)		
42	There are opportunities for students to get to know teachers outside the classroom.	.55 (.01)	.08
43	The out-of-school activities of the school have sufficient variety and scope.	.81 (.02)	.10
44	There is a good sports program in the school.	.56 (.01)	.26
53	The school places sufficient emphasis on cultural activities (music ,art, drama, etc.)	.42 (.02)	.09
	ATEC - Relationships with teachers (Chi-sq = 2770.29, df = 54, McDonald's Index = 0.80)		
56	Most teachers are well qualified and have good teaching skills.	.66 (.01)	.06
58	Most teachers in this school show a good deal of school spirit.	.66 (.01)	.10
60	Most teachers know their Year 12 students as individual persons.	.57 (.01)	.04
64	Most teachers are part of the school community.	.70 (.01)	.07
65	Most teachers have a professional attitude towards their teaching.	.68 (.01)	.03
66	Most teachers carry out their work with energy and pleasure.	.76 (.01)	.04
72	Most teachers take a personal interest in me.	.67 (.01)	.04
80	Most teachers give students sufficient encouragement.	.72 (.01)	.03
83	If students have difficulty with their school work, most teachers take time to help them.	.68 (.01)	.04
85	Most teachers go out their way to help you.	.72 (.01)	.05
86	Most teachers show that people are more important than rules.	.57 (.01)	.04
90	Most teachers here are caring and willing to assist students who need help.	.73 (.01)	.04
	STDM - Student Morale (Chi-sq = 1566.18, df = 35, McDonald's Index = 0.88)		
55	Students here think a lot of their school.	.62 (.01)	.15
69	Everyone has a lot of fun at this school.	.67 (.01)	.07
70	A good spirit of community exists amongst Year 12 students.	.62 (.01)	.07
73	I have been happy at school.	.77 (.01)	.03
76	Students at this school do not mind wearing the school uniform.	.49 (.01)	.06
79	Everyone tries to make you feel at home in this school.	.67 (.01)	.06
87	My experience of this school has been a happy one.	.82 (.01)	.03
91	I am happy to be a student at this school.	.84 (.01)	.05
92	School rules here encourage self-discipline and responsibility.	.58 (.01)	.04
93	There is a happy atmosphere in the school.	.78 (.01)	.07
	PRIN - Attitudes to the Principal (Chi-sq = 148.77, df = 5, McDonald's Index = 0.99)		
62	I can approach the Principal at any time for advice and help.	.63 (.01)	.18
71	The Principal ensures that the school provides a good education to students.	.70 (.01)	.09
81	The Principal encourages a sense of community and belonging in the school.	.81 (.01)	.15
89	The Principal places importance on the religious nature o f the Catholic school.	.42 (.01)	.10
94	The Principal provides good leadership of the school community.	.81 (.01)	.17
	ATTD - Attitudes toward Discipline (3) (Chi-sq = 24.13, df = 2, McDonald's Index = 0.998)		
59	Year 12 students are not given sufficient real freedom here.	.63 (.01)	.13
74	This school places too much emphasis on external conformity to rules and regulations.	.60 (.01)	.09
78	Most teachers never explain why they ask you to do things around here.	.42 (.01)	.04
84	There are too many rules which restrict students' freedom.	.78 (.01)	.13

Notes:

(1) Maximum Likelihood estimates using COSAN (Fraser, 1987) for the Single level Factor analysis and Maximum Likelihood estimates on intra class correlations using the ML3 package (Rasbash, Prosser, Goldstein, 1990) were reported here.

(2) Standard errors of factor loadings were printed in brackets.

(3) These items were reversely coded.

Table 2. Bilevel Factor Loadings and Single level factor loadings [1]

(a) Scale: Quality of Formal Curriculum (QUAL, 6 items, intraclass correlation = 0.08)

Item	Model A Constrained equal, No scaling	Model B Unconstrained, No scaling	Model C Constrained equal, Scaling both levels	Model D Constrained equal, Scaling at both levels, estimating sch-level factor var	Model E Unconstrained, Scaling at both levels	Single level Results
Bilevel Results : School level						
41	.68	.21	.63	.63	.94	
		(.03)	(.01)	(.01)	(.03)	
45	.70	.41	.63	.63	.85	
		(.06)	(.01)	(.01)	(.05)	
46	.64	.21	.71	.70	.96	
		(.03)	(.01)	(.01)	(.02)	
47	.63	.15	.74	.73	.95	
		(.02)	(.01)	(.01)	(.03)	
49	.68	.14	.66	.66	.89	
		(.02)	(.01)	(.01)	(.06)	
50	.65	.17	.65	.65	.87	
		(.03)	(.01)	(.01)	(.05)	
Bilevel Results : Student level						
41	.68	.68	.63	.63	.62	.64
		(.01)	(.01)	(.01)	(.01)	(.01)
45	.70	.70	.63	.63	.63	.64
		(.01)	(.01)	(.01)	(.01)	(.01)
46	.64	.63	.71	.70	.70	.72
		(.01)	(.01)	(.01)	(.01)	(.01)
47	.63	.63	.74	.73	.73	.74
		(.01)	(.01)	(.01)	(.01)	(.01)
49	.68	.68	.66	.66	.66	.66
		(.01)	(.01)	(.01)	(.01)	(.01)
50	.65	.65	.65	.65	.65	.66
		(.01)	(.01)	(.01)	(.01)	(.01)
Goodness of fit of models						
Chi-sq (Unbalanced run)	757.1	721.8	769.3	727.6	721.8	755.8
Chi-sq (Pseudo-balanced)	766.6	728.3	780.5	734.7	721.3	
d.f.	23	18	24	23	18	9
Estimates of proportion of variation at School-level						
Estimate	proportion of School-level factor variance = 0.07	$\lambda'_B\lambda_B / (\lambda'_B\lambda_B + \lambda'_W\lambda_W) =$ 0.11	Not Applicable	proportion of School-level factor variance = 0.64	$\alpha'_B\alpha_B / (\alpha'_B\alpha_B + \alpha'_W\alpha_W) =$ 0.65	

Notes:
(1) Single level factor results were repeated here to facilitate comparisons between models

Table 2. Bilevel Factor Loadings and Single level factor loadings (Cont.) (1)

(b) Scale: Out-of-School Curriculum (OUTS, 4 items, Intraclass correlation = 0.16)

Item	Model A Constrained equal, No scaling	Model B Unconstrained, No scaling	Model C Constrained equal, Scaling both levels	Model D Constrained equal, Scaling at both levels, estimating sch-level factor var	Model E Unconstrained, Scaling at both levels	Single Level Results
Bilevel Results : School level						
42	.64	.25	.56	.56	.73	
	(.02)	(.05)	(.01)	(.01)	(.08)	
43	.91	.42	.78	.78	1.09	
	(.02)	(.05)	(.01)	(.01)	(.07)	
44	.63	.46	.54	.54	.66	
	(.02)	(.09)	(.01)	(.01)	(.09)	
53	.49	.12	.42	.41	.31	
	(.02)	(.06)	(.01)	(.01)	(.14)	
Bilevel Results : Student Level						
42	.64	.64	.56	.56	.56	.55
	(.02)	(.02)	(.01)	(.01)	(.01)	(.01)
43	.91	.90	.78	.78	.78	.81
	(.02)	(.02)	(.01)	(.01)	(.01)	(.02)
44	.63	.63	.54	.54	.54	.56
	(.02)	(.02)	(.01)	(.01)	(.01)	(.01)
53	.49	.49	.42	.41	.42	.42
	(.02)	(.02)	(.01)	(.01)	(.01)	(.02)
Goodness of fit of models						
Chi-sq (Unbalanced run)	46.8	35.5	58.4	39.4	35.5	68.1
Chi-sq (Pseudo-balanced)	49.3	39.3	59.3	42.5	39.3	
d.f.	23	18	24	23	18	9
Estimates of proportion of variation at School-level						
Estimate	proportion of School-level factor variance = **0.17**	$\lambda'_B\lambda_B / (\lambda'_B\lambda_B + \lambda'_W\lambda_W)$ = **0.20**	Not Applicable	proportion of School-level factor variance = **0.64**	$\alpha'_B\alpha_B / (\alpha'_B\alpha_B + \alpha'_W\alpha_W)$ = **0.62**	

Table 2. Bilevel Factor Loadings and Single level factor loadings (Cont.) (1)

(c) Scale: Attitudes towards the Principal (PRIN, 5 items) Intraclass correlation = 0.18)

Item	Model A Constrained equal, No scaling	Model B Unconstrained, No scaling	Model C Constrained equal, Scaling both levels	Model D Constrained equal, Scaling at both levels, estimating sch-level factor var	Model E Unconstrained, Scaling at both levels	Single Level Results
Bilevel Results : School level						
62	.76	.42	.61	.60	.73	
	(.02)	(.07)	(.01)	(.01)	(.07)	
71	.68	.28	.68	.68	.90	
	(.01)	(.04)	(.01)	(.01)	(.03)	
81	.85	.45	.78	.78	.98	
	(.01)	(.05)	(.01)	(.01)	(.02)	
89	.40	.14	.42	.42	.43	
	(.01)	(.05)	(.01)	(.01)	(.13)	
94	.90	.50	.78	.77	.95	
	(.01)	(.06)	(.01)	(.01)	(.02)	
Bilevel Results : Student Level						
62	.76	.76	.61	.60	.61	.63
	(.02)	(.02)	(.01)	(.01)	(.01)	(.01)
71	.68	.68	.68	.68	.68	.70
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
81	.85	.85	.78	.78	.77	.81
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
89	.40	.40	.42	.42	.42	.43
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
94	.90	.89	.78	.77	.78	.81
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
Goodness of fit of models						
Chi-sq (Unbalanced run)	125.6	112.6	162.4	117.5	112.6	148.8
Chi-sq (Pseudo - balanced)	130.9	119.9	174.4	123.9	119.4	
d.f.	23	18	24	23	18	9
Estimates of proportion of variation at School-level						
Estimate	proportion of School-level factor variance = 0.21	$\lambda'_B\lambda_B/(\lambda'_B\lambda_B + \lambda'_W\lambda_W)$ = 0.21	Not Applicable	proportion of School-level factor variance = 0.61	$\alpha'_B\alpha_B/(\alpha'_B\alpha_B + \alpha'_W\alpha_W)$ = 0.21	

Table 2. Bilevel Factor Loadings and Single level factor loadings (Cont.) (1)

(d) Scale : Attitudes toward Discipline (ATTD, 4 items, Intraclass correlation = 0.16)

Item	Model A Constrained equal, No scaling	Model B Unconstrained, No scaling	Model C Constrained equal, Scaling both levels	Model D Constrained equal, Scaling at both levels, estimating sch-level factor var	Model E Unconstrained, Scaling at both levels	Single level Results
Bilevel Results : School Level						
59	.75	.46	.58	.58	.94	
	(.02)	(.06)	(.01)	(.01)	(.02)	
74	.61	.31	.57	.57	.93	
	(.02)	(.04)	(.01)	(.01)	(.03)	
78	.43	.14	.42	.42	.65	
	(.02)	(.03)	(.01)	(.01)	(.10)	
84	.87	.47	.74	.72	1.02	
	(.02)	(.05)	(.01)	(.01)	(.01)	
Bilevel Results : Student Level						
59	.75	.74	.59	.58	.58	.63
	(.02)	(.02)	(.01)	(.01)	(.01)	(.01)
74	.61	.61	.57	.57	.56	.60
	(.02)	(.02)	(.01)	(.01)	(.01)	(.01)
78	.43	.45	.42	.42	.42	.42
	(.02)	(.02)	(.01)	(.01)	(.01)	(.01)
84	.87	.87	.74	.72	.73	.78
	(.02)	(.02)	(.01)	(.01)	(.01)	(.01)
Goodness of fit of models						
Chi-sq (Unbalanced run)	41.2	23.2	117.9	32.0	23.2	24.1
Chi-sq (Pseudo - balanced)	43.0	24.3	123.9	34.1	24.3	
d.f.	23	18	24	23	18	9
Estimates of proportion of variation at School-level						
Estimate	proportion of School-level factor variance = 0.23	$\lambda'_B\lambda_B / (\lambda'_B\lambda_B + \lambda'_W\lambda_W)$ = 0.23	Not Applicable	proportion of School-level factor variance = 0.68	$\alpha'_B\alpha_B / (\alpha'_B\alpha_B + \alpha'_W\alpha_W)$ = 0.70	