Multilevel modeling of educational data with cross-classification and missing identification for units

(Hill, P. W. and Goldstein, H. (1998). Multilevel modelling of educational data with cross classification and missing identification of units. *Journal of Educational and Behavioural statistics* **23**: 117-128.)

Peter W. Hill
Centre for Applied Educational Research
The University of Melbourne

and

Harvey Goldstein
Institute of Education, University of London

Abstract

This paper presents a method for handling educational data in which students belong to more than one unit at a given level, but there is missing information on the identification of the units to which students belong. For example, a student might be classified as belonging sequentially to a particular combination of primary and secondary school, but for some students, the identity of either the primary or secondary school may be unknown. Similar situations arise in longitudinal studies in which students change school or class from one year to the next. The method involves setting up a cross-classified model, but replacing (0,1) values for unit membership with weights reflecting probabilities of unit membership in cases where membership information is randomly missing. The method is illustrated with reference to longitudinal data on students' progress in English.

Acknowledgments

We are most grateful to Ken Rowe for comments on a draft. This work was partly supported by the Economic and Social Research Council (UK) under its programme for the Analysis of Large and Complex Datasets. We would also like to thank two anonymous reviewers who provided detailed and constructive comments on an early version of this paper.

Introduction

Within the field of education and indeed across a wide range of disciplines, it is commonly the case that data have a complex hierarchical structure. Subjects may be clustered not only into hierarchically ordered units (e.g., students nested within classes, within schools), but may also belong to more than one unit at a given level of a hierarchy. For example, a student might be classified as belonging sequentially to a particular combination of primary school and secondary school, in which case the student will be identified by a cross classification of primary schools and secondary schools. Alternatively, a particular student may spend a proportion of time in one school and the remaining proportion in another school. In this case, the student has multiple membership of units at a given level of clustering.

Goldstein (1987) and Raudenbush (1993) present the general structure of a model for handling complex hierarchical structuring with random cross classifications. For example, assuming that we wish to model the achievement of students taking into account both the primary and the secondary school attended by each student, then we have a cross classified structure, which can be modeled as follows:

$$\begin{aligned} &y_{i(j_1 j_2)} = (X\beta)_{i(j_1 j_2)} + u_{j_1} + u_{j_2} + e_{i(j_1 j_2)}, \\ &j_1 = 1, \dots J_1, \quad j_2 = 1, \dots J_2, \quad i = 1, \dots N \end{aligned} \tag{1}$$

in which the score of student i, belonging to the combination of primary school j_1 and secondary school j_2, is predicted by a set of fixed coefficients $(X\beta)_{i(j_1, j_2)}$. The random part of the model is given by two level 2 residual terms, one for the primary school attended by the student (u_{j_1}) and one for the secondary school attended (u_{j_2}); and the usual level 1 residual term for each student. We note that the latter may be further modeled to produce complex level 1 variation (Goldstein, 1995, Chapter 3).

Rasbash and Goldstein (1994) give details of a method for estimating cross-classified models using a simple hierarchical formulation and a set of (0,1) dummy variables for each unit of one of the cross-classified random variables. The dummy variables are introduced as explanatory variables into the random part of the model and the variances of the random coefficients of these dummy variables are constrained to be equal, thus providing an estimate of the between unit variance. The method can be used to analyze a wide variety of models with the only serious limitation being the computational demands generated by models with a large number of cells in the cross classification. Examples of several kinds of frequently occurring situations in which it may be appropriate to use a multilevel random cross classification model are given by Goldstein (1995).

This paper extends the notion of the random cross classification and of multiple unit membership to encompass situations in which there is incomplete information on the units to which students belong. In presenting the rationale for this approach, however, we first consider the case where students have multiple membership of a group at a given level and we wish to assign weights other than zero or one to indicate their membership of each unit within the group.

For example, suppose that we know, for each individual, the weight π_{ij_2}, associated with the j_2-th secondary school for student i (for example the proportion of time spent in that school) with $\sum_{j_2=1}^{J_2} \pi_{ij_2} = 1$. Note that we allow the possibility that for some (typically most) students only one school is involved so that one of these probabilities is one and the remainder are zero. We can now rewrite [1] as follows:

$$
\begin{aligned}
&y_{i(j_1,j_2)} = (X\beta)_{i(j_1,j_2)} + u_{j_1}^{(1)} + \sum_{j_2} u_{j_2}^{(2)} \pi_{ij_2} + e_{i(j_1,j_2)} \\
&\sum_{j_2} u_{j_2}^{(2)} \pi_{ij_2} = \pi_i u^{(2)} \\
&u^{(2)^T} = \{u_1^{(2)}, \ldots\ldots u_{J_2}^{(2)}\} \qquad [2] \\
&\pi = \{\pi_1, \ldots \pi_{J_2}\} \\
&\pi_{j_2}^T = \{\pi_{1j_2}, \ldots \pi_{Nj_2}\}
\end{aligned}
$$

where N is the total number of students and $u^{(2)}$ is the $J_2 \times 1$ vector of secondary school effects. Thus [2] is a 2-level model where the level 2 variation among secondary schools is modeled using the J_2 sets of weights for subject i $(\pi_1, \ldots \pi_{J_2})$ as explanatory variables, with π_{j_2} the N x 1 vector of student weights for the j_2th secondary school. We have

$$
\begin{aligned}
&\mathrm{var}(u_{j_2}^{(2)}) = \sigma_{u2}^2, \quad \mathrm{cov}(u_{j_1}^{(1)} u_{j_2}^{(2)}) = 0 \\
&\mathrm{var}(\sum_{j_2} u_{j_2}^{(2)} \pi_{ij_2}) = \sigma_{u2}^2 \sum_{j_2} \pi_{ij_2}^2
\end{aligned} \qquad [3]
$$

Using proportions or probabilities other than (0,1) to indicate multiple unit membership, provides the basis of a method for handling missing unit identification information in complex hierarchical models. For example, we may have complete information about the secondary school attended by each student, but incomplete information about the primary school from which they came. On the other hand, knowing the locality of the secondary school they currently attend, we may have a reasonable basis for assigning probabilities for attendance at one or more identifiable primary schools.

Thus, we now consider the case where π_{ij_1} represents the weight associated with membership of primary school j_1 for student *i*. In many applications this will simply be the posterior probability of belonging to school j_1 and will generally depend on school size and sample design. We shall also assume that where the actual membership is unknown the student does in fact belong to just one primary school.[1] For simplicity we assume known

---

1 If we knew that some students did in fact belong to more than one school then a weighting system as in [2] would need to be used in addition to the following, leading to new weights as a product of the two.

membership of just one secondary school for each student. Although we do not know the primary school membership, the level 2 contribution to the variance is still σ_{u1}^2. Since we are ignoring the secondary school weights, this implies that [2] becomes

$$
\begin{aligned}
&y_{i(j_1,j_2)} = (X\beta)_{i(j_1,j_2)} + \sum_{j_1} u_{j_1}\sqrt{\pi_{ij_1}} + u_{j_2} + e_{i(j_1,j_2)} \\
&\mathrm{var}(\sum_{j_1} u_{j_1}\sqrt{\pi_{ij_1}}) = \sigma_{u1}^2 \sum_{j_1} (\sqrt{\pi_{ij_1}})^2 = \sigma_{u1}^2 \qquad [4] \\
&\sum_{j_1} \pi_{ij_1} = 1
\end{aligned}
$$

In the special case where we assume $\pi_{ij_1} = \sqrt{1/n_{j_1}}$, where n_{j_1} is the number of students in the j_1 primary school, representing complete agnosticism about primary school membership, this leads to the intra-primary school correlation between two students who are actually in the same primary school, but neither of whose primary school membership is known, becoming

$$
\left(\frac{\sigma_{u1}^2}{\sigma_e^2 + \sigma_{u1}^2}\right) / n_{j_1} \qquad [5]
$$

which reflects the fact that the probability of two randomly chosen students belonging to the same school is $1/n_{j_1}$. Finally we can combine (2) and (4) to deal with both missing identifications and multiple unit membership.[2]

A worked example

To illustrate the application of the above methodology, we make use of data from a three-year longitudinal study of educational effectiveness known as the Victorian Quality Schools Project (Hill, Holmes-Smith, & Rowe, 1993; Hill & Rowe, 1996).

A two-stage, stratified, probability sample of government, Catholic and independent schools in the State of Victoria, Australia, was drawn on the basis of an estimated intra-unit correlation of 0.2 and an average cluster size of 30 (see Ross, 1988a, 1988b). Within these constraints, schools were randomly selected at the first stage of sampling, but with probability proportional to their enrollment size. At the second stage of sampling, the total

---

[2] Note added after publication: Equations (2) and (4) do not provide a consistent notation, although the intent is clear. A consistent general notation for a 2-level data structure is as follows:

$$
y_{i\{j\}} = (X\beta)_{i\{j\}} + \sum_{h\in\{j\}} u_h \pi_{ih} + e_{i\{j\}}
$$

where *j* indexes the *m* level 2 units and the level 1 units are indexed uniquely by *i*. The weight π_{ih} is that associated with the *h*-th level 2 unit for the *i*-th level 1 unit. For the particular cases of a simple hierarchy where there is just one level 2 unit for each level 1 unit, or a 2-way cross classification where there are just two such units then $y_{i\{j\}}$ can be replaced respectively by y_{ij}, $y_{i(j_1j_2)}$ without ambiguity.

cohort of students enrolled in the Kindergarten or Preparatory Grade (K), Grade 2, Grade 4, Grade 7 and Grade 9 in each selected school, were included in the sample.

For illustrative purposes, we focus on data relating to teacher ratings of the achievement in English of primary school students. The English scores have been scaled to have a mean of zero and a standard deviation of 1. In the first year of the study (1992), useable data were received from 59 primary schools including 41 government schools, 12 Catholic schools and 6 independent schools, for a total of 6,678 students and 365 teachers. In the second and third years of the study, data were obtained on the same students remaining in the sampled schools as they proceeded to Grades 1, 3 and 5 and 2, 4 and 6 respectively, as indicated diagrammatically below .

1992	**K**		**2**		**4**		
1993	↘	**1**	↘	**3**	↘	**5**	
1994		↘	**2**	↘	**4**	↘	

Sample attrition rates over the life of the project were relatively high due to a number of factors, one of which was missing data arising from a failure on the part of respondents to answer all questions. However, natural attrition also played a part. Australia is a highly mobile society and a high turnover of students from year to year is common. In addition, policy changes saw the closure of around one in ten Government schools over the three-year period during which the study was in progress and this also had an impact on sample attrition. Finally, five schools dropped out of the project after the first year for a variety of reasons, but mainly on account of workload pressures on teachers. In most cases, the missing data could be regarded as effectively randomly missing rather than systematically related to the characteristics of the students retained in the sample.

In modeling the English achievement of students we assume a bivariate response model of the general form

$$\begin{aligned} y_{ti(j_1,j_2 j_3)k} &= (X\beta)_{ti(j_1,j_2,j_3)k} + \beta_z z_{ti(j_1,j_2,j_3)k} \\ &\quad + (v_{tk} + u^{(1)}_{tj_1} + u^{(2)}_{tj_2} + u^{(3)}_{j_3} + e_{ti(j_1,j_2,j_3)k}) \end{aligned} \qquad [6]$$

in which the achievement of student *i*, either at t=1, namely at the end of 1993, or at t=2, namely at the end of 1994, is predicted by a set of student characteristic variables $X_{ti(j_1,j_2,j_3)}$, which may be background or time-varying measures, and a measure of prior achievement $z_{ti(j_1,j_2,j_3)}$ taken at the end of 1992 when t=1 and at the end of 1993 when t=2. The class level terms $u^{(1)}_{tj_1}, u^{(2)}_{tj2}$ and $u^{(3)}_{j3}$ respectively refer to the 1992, 1993 and 1994 class membership of the students. The subscript k indexes schools and v_{tk} is the effect of school k at time t, and $e_{ti(j_1,j_2,j_3)t}$ is the contribution of student i within school k in classes

j_1(1992 class), j_2 (1993 class) and j_3(1994 class) at time t. The subscript t is thus the indicator for the response. We have not incorporated a subscript t for the 1994 class, since we assume that membership of the 1994 class will affect only the 1994 response. By contrast for the 1992 and 1993 class we have a residual term for both the 1993 response and the 1994 response. Likewise both responses are present at the student and school levels. We also note that the response in 1993 is the prior achievement covariate for the 1994 response. The model [6] can be represented as a five-level model, with observations at t=1 and t=2 nested within students within classes within schools. To accommodate the fact that class composition changed in each of 1992, 1993 and 1994, a cross classification of 1993 classes with 1992 and 1994 classes is introduced into the model by declaring two additional levels in the model, making it a six-level model (see Appendix). Note that no level 1 random terms appear in the model, since this defines the bivariate structure (Goldstein, 1995, Chapter 4). Thus, in the random part of the model, there are five variables representing student 1992 class, 1993 class, 1994 class and school effects.
Equation [6] assumes that for each student there are two records or sets of observations, one relating to achievement at the end of 1993 and including a 1992 prior achievement measure, and the other relating to achievement at the end of 1994 and including a 1993 prior achievement measure. For each 1993 record, it is assumed that unit identification includes both the 1992 and 1993 class to which each student was assigned and the school to which each student belongs. For each 1994 record, it is assumed that unit identification includes the 1992, 1993 and 1994 class to which each student was assigned and the school identification.

In practice, it was found that there were 460 students with a missing 1992 class identification. To have carried out an analysis using only cases where there was complete identification would have been to reduce efficiency. Also, if missingness is not completely random, by including all available data we will tend to reduce any possible biases.

To incorporate all (4539) students and (6423) records within the one analysis, the assumption was made that the 460 students with missing 1992 class identification information had an equal probability of belonging to any one of the classes within the same school in 1992. Accordingly, we used identification weights calculated as $\sqrt{1/n_{j_2}}$. We note that our procedure is designed to deal with missing *identification* in 1992 records rather than complete missing 1994 records. In the latter case, assuming missingness at random, the data set will be unbalanced with respect to time but we will still obtain efficient (maximum likelihood) estimates (Goldstein, 1995, Chapter 4).

Table 1 summarizes the results of fitting [6] to the data using the above method for handling missing unit identification information within the model. Model parameters were estimated using the multilevel software *MLn* (Rasbash, Goldstein, & Woodhouse, 1995). A description of how the model was implemented using *MLn* is provided in the Appendix.

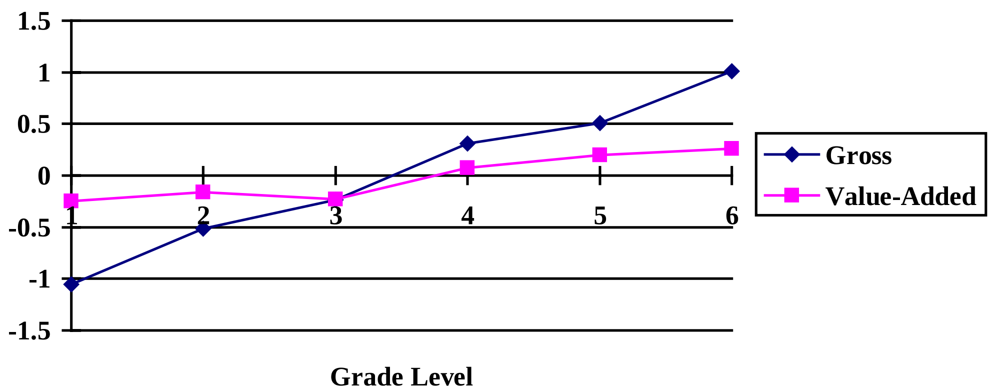
Table 1. Parameter estimates (and Standard Errors) for two bivariate response cross-classified models with missing identification codes. Model 2b is identical to model 2a but with subjects excluded who have a missing 1992 class identification.

Parameters	Model 1	Model 2a	Model 2b

	(N=6423)	(N=6423)	(N=5963)
Fixed:			
1993 intercept (t=1)	-1.054 (0.010)	-0.246 (0.049)	-0.223 (0.051)
Grade 3 (t=1)	0.815 (0.057)	0.019 (0.063)	0.103 (0.064)
Grade 5 (t=1)	1.561 (0.055)	0.444 (0.062)	0.342 (0.066)
1994 intercept (t=2)	-0.515 (0.051)	-0.164 (0.051)	-0.032 (0.056)
Grade 4 (t=2)	0.823 (0.064)	0.236 (0.069)	0.061 (0.078)
Grade 6 (t=2)	1.527 (0.066)	0.422 (0.071)	0.365 (0.083)
Gender (Female)	-	0.067 (0.010)	0.056 (0.010)
Non-English Speaking	-	-0.045 (0.020)	-0.028 (0.021)
Occupational Status	-	0.059 (0.006)	0.058 (0.007)
Critical Events	-	-0.045 (0.012)	-0.036 (0.012)
Prior Achievement	-	0.700 (0.010)	0.721 (0.010)
Random:			
School			
σ^2_{v1} (1994)	0.052 (0.019)	0.027 (0.016)	0.008 (0.014)
σ^2_{v0} (1993)	0.018 (0.010)	0.044 (0.017)	0.032 (0.014)
σ_{v10} (1993,1994)	0.026 (0.011)	0.010 (0.012)	0.012 (0.010)
Class			
$\sigma^2_{u(j3)0}$ (1994)	0.181 (0.020)	0.184 (0.019)	0.221 (0.024)
$\sigma^2_{u(j2)1}$ ('93 class for '94 response)	0.012 (0.006)	0.095 (0.011)	0.102 (0.013)
$\sigma^2_{u(j2)0}$ ('93 class for '93 response)	0.175 (0.017)	0.176 (0.017)	0.145 (0.015)
$\sigma_{u(j2)10}$ (1993,1994)	0.004 (0.008)	-0.129 (0.013)	-0.119 (0.013)
$\sigma^2_{u(j1)0}$ ('92 class for '94 response)	0.002 (0.004)	0.002 (0.003)	0.001 (0.002)
$\sigma^2_{u(j1)1}$ ('92 class for '93 response)	0.009 (0.004)	0.085 (0.010)	0.102 (0.011)
Student			
σ^2_{e1} (1994)	0.279 (0.008)	0.132 (0.004)	0.130 (0.004)
σ^2_{e0} (1993)	0.281 (0.007)	0.151 (0.004)	0.150 (0.004)
σ_{e10} (1994,1993)	0.203 (0.006)	-0.030 (0.004)	-0.035 (0.004)
-2*log(likelihood)	10494	8149	
Intra-school correlations			
1994 responses	0.094	0.061	
1993 responses	0.037	0.096	
1994 + 1993 responses	0.069	0.079	
Intra-class correlations			
1994 responses (1994 classes)	0.330	0.418	
1993 responses (1993 classes)	0.353	0.386	
1994 + 1993 responses	0.376	0.606	
σ_{e10} (1993,1994)	0.197 (0.006)	-0.029 (0.004)	-0.039 (0.004)
-2*log(likelihood)	107	8289	

Considering first the fixed parameter estimates, the 'intercept' variables '1993' and '1994' are the achievement levels, expressed in standardised units and adjusted for all other explanatory variables in the model, of students in the earliest years of schooling, namely Grade 1 and Grade 2, in 1993 and 1994. Then follow dummy variables for the other year levels, namely Grades 3 to 6, so that the associated coefficients represent the differences from the Grade 1 and Grade 2 values. Model 1 predicts achievement scores solely on the basis of Grade level, thus providing a 'base' model with which to evaluate the effects of including further explanatory variables. By adding the coefficients for each Grade level to the coefficients for the base Grades (Grade 1 in 1993 and Grade 2 in 1994), estimates can be obtained of average gross achievement levels and (from model 2) of average adjusted or 'value-added' achievement levels across Grades 1 to 6.[3] These are graphed in Figure 1.

Figure 1. Gross and value-added average scores for Grades 1 to 6



Turning now to the random part of Model 1, it will be noted that the residual variation at the school level is is small and higher for 1994 responses than for 1993 responses, indicating a tendency for the gap between schools to grow. The total residual variance for the 1993 response is 0.48 and for the 1994 response is similar at 0.53. The proportion of variance attributable to between school differences is about four per cent for 1993 responses and nine per cent for 1994 responses. It will be noted also from the covariance terms ($\sigma_{v10}, \sigma_{e10}$) that there is a strong, positive correlation at the school level between unadjusted 1993 and 1994 achievement scores (r = 0.849) so that considered jointly, the intra-school correlation is 0.069.

The parameter estimates for residual variance at the class level (levels 3, 4 and 5) of Model 1 indicate that if 1993 and 1994 responses are considered separately, there is an intra-class correlation of 0.353 and 0.330 respectively, indicating substantial differences among classes

3 A more elaborate model which allowed interactions between year and other explanatory variables, that is separate coefficients for each response, was also fitted. The differences in the coefficients between 1993 and 1994 were small, however, and only the simpler model is represented.

in teacher ratings of student achievement. Adding 1993 and 1994 effects, this translates into an overall intra-class correlation of 0.376. As one might expect, the effect of 1993 class membership on 1993 responses is significantly greater than its effect on 1994 responses, but the magnitude of the effect on 1994 responses is nevertheless significant.

In Model 2a, student achievement in English is predicted not only by Grade level, but also by four student background characteristics and by Prior Achievement measured one year previously. The parameter estimates indicate that female students make greater progress than males and students from high occupational status families make greater progress than students from families where the highest breadwinners are unemployed or in unskilled occupations. Non-English Speaking Background students make less progress than their English-speaking counterparts and having experienced a Critical Event during the year, such as an extended illness or some psychological trauma, is associated with a negative effect on student progress. Prior Achievement is by far the most important predictor of student progress.

Inclusion of the student background variables and the measure of Prior Achievement result in a significant improvement in model fit as indicated by the log likelihood ratio. Looking at the total (1993 + 1994) effects we note that judged one year at a time, the evidence points to large within-school-between-class differences. The largest effects of 1993 class membership are on 1993 responses, while the largest effects on 1994 class membership are on 1994 responses, however, 1993 class membership has a significant effect on 1994 responses. The effect of 1992 class membership on 1993 and 1994 responses is small and insignificant with respect to the effect on 1994 responses.

Model 2b is identical to Model 2a, but fitted to the data for only those students with complete class identification information. In other words, the 460 students with missing 1992 class identification information were omitted from the analysis. This was done to investigate any biasing effects from excluding these students from the analysis. The parameter estimates for both the fixed and the random parts of the model are in most instances very similar, suggesting that students with missing 1992 class information were broadly representative of the full sample, although there are differences between the intercepts for the different year levels.

Discussion

The above approach to handling missing identification information can be generalized to a wide range of data analytic problems and practical situations. It can be used in both univariate and multivariate linear and non-linear models involving complex hierarchical structures. Its main application, however, would appear to be in connection with longitudinal studies where problems of missing data related to the classification of students by classes or schools are frequently encountered as students change school or the composition of classes changes from one year to the next. This is particularly important for school effectiveness studies where the standard approach has been simply to omit students who change school during the course of a study, even when this involves a majority of the students. As we have pointed out, such a practice is not only inefficient, but also likely to result in biases which the present procedure can help to avoid. Although we have used a bivariate response model, the procedure can be used for a traditional repeated measures design where the same measurement is made at several occasions. The paper also presents a model where there is multiple membership of groups so that, for example, measures of ‘degree of group membership’ are required. This model can be combined with the missing identification model, as referred to in the text, so leading to a very general class of models.

As mentioned previously, the main computing limitations derive from the large amounts of storage required for model estimation. The computational complexities associated with the approach when the number of ‘identification failures’ or multiple memberships becomes large should influence study design in ways which minimize resulting problems. This is an area for further investigation, but there are certain general points which can be made. It should be noted, however, that with current rapid advances in computer hardware, these problems will become far less severe.

If students are to be followed in the same set of schools over time, it may be advisable to sample schools which are geographically strongly clustered so that moves will tend to be to schools already sampled. In a cross classified design where, say, students move from primary to secondary schools, the sampling of primary schools should be such that as many as possible of all those ‘feeding’ to a well defined set of secondary schools are chosen. If there is knowledge of mobility patterns then these should inform the sampling.

If student progress is to be monitored, longitudinal research is essential. However, researchers have been acutely aware of the difficulties with such research, since students regularly move from school to school and from class to class. Tracking these movements is often difficult and for many students there will be missing information on their group membership. The above approach to dealing with these situations extends the analytic tools available for studying the complex reality of educational settings.

Appendix

Model [6] is readily specified in *MLn* (Rasbash and Woodhouse, 1995) using the standard approach to organizing data for a bivariate model, with data sorted by school, 1993 class, student and occasion.

To create the cross classification between 1993 classes and 1992 and 1994 classes, each class was numbered sequentially within each school for each way of the cross classification (i.e., 1992, 1993 and 1994 classes). In the case of the data used in the worked example, the maximum number of classes sampled in any one school was 17 in 1992 and 1993 and 14 in 1994. Thus the class identification codes ranged from 1 to 17 for 1992 and 1993 classes, and from 1 to 14 for 1994 classes. In the case of records with missing class identification codes, it was found convenient to assign these records to class 1 and to subsequently replace this temporary identification with zeros through the use of the SETX command.

1993 class membership was handled in the normal way at level 3 in the model. It will be recalled that the membership of 1993 and 1994 classes was known for all students. The SETX command was used to generate a set of dummy variables for 1992 and 1994 class membership, assigning the variables to levels 4 and 5 of the model. The explanatory variable used to generate the 14 dummy variables for 1994 class membership was a dummy variable coded '1' if the student's 1994 class identification was known and '0' if otherwise. A similar process was used for 1992 class membership.

The next step was to replace zeros with an appropriate set of weights for those students whose 1992 class membership was unknown. For example, suppose we had three records as shown below. The first two records relate to student 1 who was in class 4 in school *k* in 1993 and class 3 in school *k* in 1992. The third record relates to student 3 who was in class 4 in 1993, but whose 1992 class is unknown. Therefore, we assign a set of weights reflecting the probability of being in one of the three classes in school *k* in 1992, namely (assuming equal probabilities) $\sqrt{1/3}$ = 0.577.

Student	Record	1993 Class				1992 Class			
		1	2	3	4	1	2	3	4
1	1	0	0	0	1	0	0	1	0
1	2	0	0	0	1	0	0	1	0
2	1	0	0	0	1	0.577	0.577	0.577	0

There are several ways of inserting the relevant weights into the set of dummy variables representing class membership. It was found that the most efficient way of doing this was to use CALC commands within a loop.

To enable separate estimates for the effect of 1992 class membership on 1993 and 1994 responses, separate dummy variables were generated for the 17 classes for both 1993 and 1994 responses.

The final step was to ensure that the appropriate constraint matrix was generated and attached to the model to constrain each of the variance estimates associated with classes for a given response to be equal. Instructions relating to specifying a cross classified model with the appropriate constraints included are contained in Appendix F of the *MLn Command Reference* (Rasbash and Woodhouse, 1995). A set of Mln macros has been written to simplify the specification of models of the kind referred to in this paper and are available from the second author. In the Microsoft windows release of *MLn* (*MLwiN* - March, 1998) there is a simple set of commands for such models.

References

Castles, I. (1986). *Australian standard classification of occupations: Statistical classification*. Australian Bureau of Statistics. Canberra: C. J. Thompson Government Printer.

Goldstein, H. (1987). Multilevel covariance component models. *Biometrika*,74, 430-1

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold; New York: Halsted Press.

Hill, P.W., Holmes-Smith, P., & Rowe, K.J. (1993). *School and teacher effectiveness in Victoria: Key findings from Phase 1 of the Victorian Quality Schools Project.* Melbourne: Centre for Applied Educational Research, The University of Melbourne (ERIC Clearing House, Document No. ED 367 067).

Hill, P.W. & Rowe, K.J (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement*, 7, 1-34 .

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, *19,* (4), 337-350.

Rasbash, J., & Woodhouse, G. (1995). *MLn Command Reference*. London: Institute of Education, University of London.

Raudenbush, S.W. (1993). A crossed random effects model for unbalanced data with applications in cross sectional and longitudinal research. *Journal of Educational Statistics*, 18, 321-349.

Ross, K.N. (1988a). Sampling. In J.P. Keeves (E.), Educational research, methodology and measurement: An international handbook (pp.527-537). Oxford: Pergamon Press.

Ross, K.N. (1988b). Sampling errors. In J.P. Keeves (E.), Educational research, methodology and measurement: An international handbook (pp.537-541). Oxford: Pergamon Press.