Multilevel Modelling in Large Scale Achievement Surveys

by

Harvey Goldstein
Institute of Education
University of London

Paper read to American Educational Research Association meeting, San Francisco, March 27-31 1989.

1 Introduction

Large scale Educational Achievement surveys often are designed using complex matrix sampling so that a wide topic coverage can be obtained without imposing too great a burden upon individual student respondents. While this is a potentially efficient means of reporting student achievement and factors related to it, the potential can only be released if the statistical models used in the analysis fully exploit the sample design. In this paper I shall explore methods for doing this using data from the Second International Science Survey (International Association for the Evaluation of Educational Achievement(IEA), 1988).

The analyses reported in this paper are intended to illustrate a methodology rather than to provide extensive substantive analysis of the data set. Nevertheless, although only a small number of variables is used, some substantively interesting results are obtained as a result of using an efficient analysis procedure. Furthermore, the multilevel models which form the basis for the analyses can now be used on large data sets so that full scale modelling can be undertaken.

2 The Data

Data are available for 4 countries, which I will refer to as countries A,B,C,D. They consist of scores for grade 9 [1] students on a biology test. The test items are grouped into five sets. There is a core test consisting of 10 items which is taken by all students, and four 'rotated form' tests consisting of 3, 2, 4 and 4 items respectively. These rotated forms contain more specialised items and are administered randomly among students within classrooms. Students take two of these rotated forms in addition to the core test, although in country A some students took either no rotated form or only one. The rotated form items are expected to be more difficult on average than the core test items, and this therefore needs to be taken account of in the analysis; for two reasons. Firstly, there may be an association between form taken and other student or school characteristics which could yield spurious associations between those characteristics and test scores, if not recognised. Secondly, the extra variation induced by form differences will tend to reduce the precision of parameter estimates.

There is little intrinsic interest in separate subscores for the core and rotated forms, especially since these would be based on a small number of items. Rather, interest lies in the average score over all items to which students respond, and how this is related to other factors. Such an average score can be constructed using different relative weights for the core and rotated forms, although for the present analysis an equal weighting scheme is adopted. Thus, for each student the total raw score is calculated based on the number of correct responses out of the number of items taken and then divided by that number to form an average proportion correct.

3 A Basic Model

3.1 Notation

We denote the response variable by y_{ij} where the subscript i refers to the school and the subscript j to the student within school, yielding two levels for the data set. There are 6 explanatory variables plus the constant term in the model, as follows:

1 In the case of country B this is grade 8.

1. Student gender (female coded 1, male coded 0)

2. Opportunity to learn(OTL). Each school reported for each of the 23 items the percentage of students who had encountered the topic to which it referred. The proportion of items where more than 25% had encountered the item is defined as a measure of 'opportunity to learn'. In country B this percentage was taken as more than 60%.

3. Rotated form taken. This is defined by three (0,1) dummy explanatory variables with rotated form number 4 taken as the base category.

4. The proportion of girls in the student's class.

Thus there are four student level variables, one school level variable and one student level variable aggregated to the level of the school. A simple basic model can be written as follows

$$y_{ij} = \sum_{k=0}^{6} \beta_k x_{ij} + u_i + e_{ij} \qquad (1)$$

where x_0 is the constant =1, and

$$var(u_i) = \sigma_u^2, \quad var(e_{ij}) = \sigma_e^2$$

This is a standard 2-level variance components model in which, as explained above, separate terms are fitted for the presence of each rotated form, and the test score is assumed to have a simple within and between school variance structure. In fact, because the number of items taken by each student differs, we would expect that the between student variation changes with the number of items taken. That is, it will vary with the particular combination of rotated forms taken.

In general, if p combinations of core plus rotated forms are present in the data we would require the same number of parameters to model fully the level 1 variation. In the present case this variance heterogeneity is modelled by assuming that the level 1 variance is a variance associated with the core test plus a term for each rotated form taken. This can be written as

$$\sigma_e^2 = \sigma_{e0}^2 + \sigma_{e1}^2 z_1 + \sigma_{e2}^2 z_2 + \sigma_{e3}^2 z_3 + \sigma_{e4}^2 z_4 \qquad (2)$$

where the z are dummy variables for the rotated forms. This involves the estimation of 5 parameters in the present case and can be thought of as a 'main effects' modelling of the level 1 variance structure. In countries B, C, and D, however, all students take the core and two rotated forms, so that a separate variance for the core cannot be estimated unless the design is deliberately unbalanced by omitting some rotated form responses. In country A some students took only the core or just one rotated form and for this country a separate core test variance can be estimated.

A full descripton of how such a complex level 1 variance structure may be specified is given in Goldstein (1987), and it should also be noted that this only appears to be possible using software based upon the Iterative Generalised Least Squares (IGLS) algorithm(Rasbash et al, 1988).

3.2 Data Analysis

The basic model given by (1) and (2) is further extended by supposing that the coefficients of OTL and gender vary randomly across schools. Model (1) can now be written as

$$y_{ij} = \sum_{k=0}^{6} \beta_k x_{ij} + (u_{0i} + u_{1i} x_{1ij} + u_{2i} x_{2ij}) + e_{ij} \qquad (3)$$

The results of fitting (2) with (3) for each country are given in table 1.

(table 1 here)

In the random part of the model only country A fails to show any variation in the coefficients of gender or OTL at level 2. In country B there is a significant variation in the gender difference across schools, while in country D it is small and not significant.[2] In country C there is a variation in the coefficient of OTL across schools, suggesting that the importance of curriculum exposure varies from school to school. The proportion of the total variance which lies between schools varies from about 3% in country C (at the mean OTL score) to about 18% for girls in country B. Although the same test items were given in all countries, the approach of studying relative variation between schools extends to the case where different tests are used. Thus, interesting comparisons of relative variation can be obtained where comparison of mean scores is impossible or inappropriate.

In the fixed part of the model the coefficient of OTL only appears significant and sizeable in countries B and C. The difference between girls and boys is negative in all countries and in all countries forms 1 - 3 predict higher scores than form 4. In country A the differences between the rotated forms and the core range from 0.08 to 0.13.In country A and country D there were single gender schools, but only in the latter case were these associated significantly associated with increased scores.

The most discrepant results are related to the effect of the proportion of girls in the student's own class. In country A and country C this had only a small effect, but larger effects in countries B and D. Figure 1 shows the predicted effect of this class level variable over the range of values common to all countries. Country B is an industrialised country whereas country D is a developing country and this may be related to the differences, but further work needs to be undertaken. In particular, there is no information about thye stability of classes over years, nor is there information about the overall proportion of girls in the school.

(figure 1 here)

4 Discussion

The principal purpose of this paper has been to elaborate a methodology for the analysis of complex matrix sample designs in comparative studies of educational achievement. The emphasis has been on developing a detailed description of the variance structure at the student level so that the full potential efficiency of the design can be realised. In addition, we have studied possible variation in the linear model coefficients across schools. The IEA data set is, of course, very much richer than the few variables used in the present analysis and further work would concentrate on a study of those factors which might explain more of the variation at both the student and school level.

---

2 The standard errors associated with the random parameters are approximate and significance tests and confidence intervals should be interpreted with caution. (See Goldstein, 1987).

An issue which has not been touched on is that of efficient design. We see that the design adopted in countries B, C, and D did not allow separate estimation of the effect of the core test. Furthermore, for efficient estimation of the between-school variation a relatively large number of schools is required and a cost effective design would need to balance the estimation efficiency associated with more schools against the increased cost of obtaining such data.

A further important extension of the present analysis is to model all four countries together. In such a model the fixed coefficients could be studied for differences between countries and likewise for the random coefficients.

Finally, I have here only been concerned with a single response. In practice it is most efficient to analyse all the responses simultaneously, since the matrix sample design will extend generally to selection of topic areas covered as well as forms within topic area. The gains in efficiency from such analyses will often be very large indeed and they will also allow the estimation of the relationships between the different response variables, from different topic areas. To carry out such a multivariate analysis involves extending the current model to a 3 level model where the different topics are nested within students at level 1. A description is given in Goldstein (1987, chapter 5).

5 Acknowledgements

I am most grateful to Richard Wolfe, David Wiley and Neville Postlethwaite for their support and comments. The work was partly supported by a grant from the U.S. National Science Foundation and partly by a grant from the U.K. Economic and Social Research Council to the Multilevel Models Project at the Institute of Education.

6 Summary

Using data from the International Educational Achievement Second Science survey, an exposition is given of efficient approaches to the analysis of multiple matrix designs. It is shown how a multilevel model approach exploits the design in an efficient manner and also extends the modelling possibilities to give greater insight into educational processes.

7 References

Goldstein, H. (1987). Multilevel models in educational and social research, :London, Griffin; New York, Oxford University Press,

International Association for the Evaluation of Educational Achievement. (1988). Science Achievement in Seventeen Countries, Oxford:Pergamon Press,

Rasbash, J., Prosser, R. and Goldstein, H. (1988). ML2: Software for two level analysis, London:Multilevel Models Project, Institute of Education,

TABLE 1

Parameter Estimates for IEA Biology scores for four countries

	Country A value(s.e.)	B value(s.e.)	C value(s.e.)	D value(s.e.)
Fixed parameters.				
Constant	0.43	0.78	0.59	0.34
Gender	-0.03(0.007)	-0.01(0.006)	-0.03(0.004)	-0.03(0.005)
% girls	-0.07(0.04)	-1.00(0.56)	-0.05(0.07)	1.03(0.39)
$(\%\text{ girls})^2$	-	1.13(0.53)	-	-0.90(0.39)
Boys school	0.001(0.03)	-	-	0.33(0.10)
Girls school	0.01(0.03)	-	-	0.33(0.10)
OTL	-0.02(0.02)	0.11(0.03)	0.06(0.02)	-0.03(0.03)
Form 1	0.13(0.01)	0.02(0.01)	0.04(0.005)	0.003(0.005)
Form 2	0.13(0.01)	0.04(0.01)	0.03(0.005)	0.02(0.005)
Form 3	0.10(0.01)	0.02(0.01)	0.03(0.005)	0.004(0.005)
Form 4	0.08(0.01)	-	-	-

/ctd.

/CTD.

TABLE 1

Random Parameters.

Level 1				
variances:				
Core	0.029(0.003)	-	-	-
Form 1	-0.002(0.002)	0.004(0.006)	-0.0012(0.0014)	-0.001(0.001)
Form 2	-0.003(0.002)	0.006(0.014)	-0.0012(0.0014)	0.001(0.001)
Form 3	-0.001(0.002)	0.0002(0.001)	0.0015(0.0014)	0.000(0.001)
Form 4	-0.003(0.002)	0.017(0.002)	0.035(0.002)	0.016(0.002)
Level 2				
Cons./Cons.	0.003(0.0005)	0.005(0.001)	0.006(0.005)	0.003(0.001)
Gender/ Gender	-	0.003(0.001)	-	0.0005(0.0004)
Otl/Otl	-	-	0.014(0.011)	-
Cons./Gender	-	-0.001(0.001)	-	-0.0006(0.0004)
Cons./Otl	-	-	-0.009(0.008)	-
Gender/Otl	-	-	-	-
Number of schools	146	99	199	96
Number of students	3035	2439	7610	3780

Note: At level 2 the random parameters are the variances associated with the the coefficient of the constant(intercept), gender and OTL, together with the covariances between these three coefficients.

For countries B,C,D a separate core variance at level 1 and a fixed coefficient estimate for the core cannot be estimated; see text. In these cases the fixed coefficients measure deviations from the form 4 coefficient value. For the random parameters, form 4 actually estimates the variance of the core plus form 4, and forms 1 - 3 estimate the differences from this value for these other forms. In country A forms 1 - 4 in the fixed part estimate the additional effect of these forms over the core and in the random part the additional variance at level 1. Where estimated coefficients are small they have been excluded from the analysis which is presented.

Figure 1

Effects of Proportion of Girls in Classroom for four Countries