International Education Journal, 2006, 7(2), 160-173.
ISSN 1443-1475 © 2006 *Shannon Research Press*.
http://iej.com.au

Suppressor variables and multilevel mixture modelling

I Gusti Ngurah Darmawan
School of Education, University of Adelaide *igusti.darmawan@adelaide.edu.au*

John P. Keeves
School of Education, Flinders University *john.keeves@flinders.edu.au*

A major issue in educational research involves taking into consideration the multilevel nature of the data. Since the late 1980s, attempts have been made to model social science data that conform to a nested structure. Among other models, two-level structural equation modelling or two-level path modelling and hierarchical linear modelling are two of the techniques that are commonly employed in analysing multilevel data. Despite their advantages, the two-level path models do not include the estimation of cross-level interaction effects and hierarchical linear models are not designed to take into consideration the indirect effects. In addition, hierarchical linear models might also suffer from multicollinearity that exists among the predictor variables. This paper seeks to investigate other possible models, namely the use of latent constructs, indirect paths, random slopes and random intercepts in a hierarchical model.

Multilevel data analysis, suppressor variables, multilevel mixture modelling, hierarchical linear modelling, two-level path modelling

INTRODUCTION

In social and behavioural science research, data structures are commonly hierarchical in nature, where there are variables describing individuals at one level of observation and groups or social organisations at one or more higher levels of observation. In educational research, for example, it is interesting to examine the effects of characteristics of the school, the teacher, and the teaching as well as student characteristics on the learning or development of individual students. However, students are nested within classrooms and classrooms are nested within schools, so the data structure is inevitably hierarchical or nested.

Hierarchical data structures are exceedingly difficult to analyse properly and as yet there does not exist a fully developed method for how to analyse such data with structural equation modelling techniques (Hox, 1994, as cited in Gustafsson and Stahl, 1999). Furthermore, Gustafsson and Stahl (1999) mentioned that there are also problems in the identification of appropriate models for combining data to form meaningful and consistent composite measures for the variables under consideration.

Two commonly used approaches in modelling multilevel data are two-level structural equation modelling or two-level path modelling and hierarchical linear modelling. Despite their advantages, the two-level path models currently employed do not include the estimation of cross-level interaction effects; and hierarchical linear models are not designed to take into consideration the latent constructs as well as the indirect paths. In addition, some other problems are associated with the use of HLM, such as fixed X-variables with no errors of measurement, limited modelling possibilities and like any regression the analysis also suffers from the multicollinearity that exists among the predictor variables. The multicollinearity issue is considered in the following section because discussion of this issue is not only highly relevant, but is also rarely undertaken.

MULTICOLLINEARITY AND SUPPRESSOR VARIABLE

Since Horst (1941) introduced the concept of the 'suppressor variable', this problem has received only passing attention in the now nearly two-thirds of a century since it was first raised. In its classical rendering Conger (1974) argued that a suppressor variable was a predictor variable, that had a zero (or close to zero) correlation with the criterion, but nevertheless contributed to the predictive validity of a test.

Three types of suppressor variables have been identified. Conger (1974) labelled them as traditional, negative and reciprocal. Cohen and Cohen (1975) named the same categories classical, net, and cooperative. To describe these three types of suppression, suppose that there are the criterion variable Y and two predictor variables, X_1 and X_2.

Classical Suppression

A classical suppression occurs when a predictor variable has a zero correlation with the criterion but is highly correlated with another predictor in the regression equation. In other words, $r_{Y1} \neq 0$, $r_{Y2} = 0$, and $r_{12} \neq 0$. In order to understand the meaning of these coefficients it is useful to consider the Venn diagram shown in Figure 1.

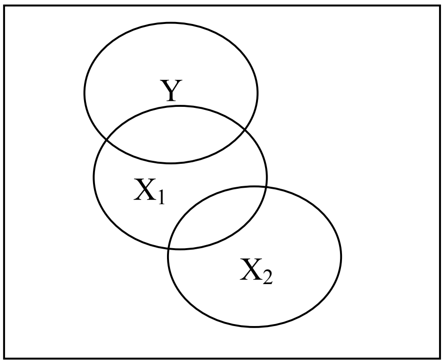

Figure 1. A Venn diagram for classical suppression

Here the presence of X_2 increases the multiple correlation (R^2), even though it is not correlated with Y. What happens is that X_2 suppresses some of what would otherwise be error variance in X_1.

Cohen et al. (2003, p.70) gave the formula for the multiple correlation coefficient for two predictors and one criterion as a function of their correlation coefficients:

$$R^2_{Y.12} = \frac{r^2_{Y1} + r^2_{Y2} - 2r_{Y1}r_{Y2}r_{12}}{1 - r^2_{12}} \tag{3}$$

Since $r_{Y2} = 0$, equation (3) can be simplified as

$$R^2_{Y.12} = \frac{r^2_{Y1}}{1 - r^2_{12}} \tag{4}$$

Because r^2_{12} must be greater than 0, the denominator is less than 1.0. That means that $R^2_{Y.12}$ must be greater than $r^2_{Y.1}$. In other words, even though X_2 is not correlated with Y, having it in the equation raises the R^2 from what it would have been with just X_1. The general idea is that there is some kind of noise (error) in X_1 that is not correlated with Y, but is correlated with X_2. By including X_2 this noise is suppressed (accounted for) leaving X_1 as an improved predictor of Y. The magnitude of the $R^2_{Y.12}$ depends of the values of r_{12} and r_{1Y} as can be seen in Figure 2, where the multiple correlation ($R^2_{Y.12}$) for different values of r_{12} and for the different correlations between X_1 and Y have been presented. In some cases, the $R^2_{Y.12}$ value can be greater than 1.

Cohen et al. (2003, p. 68) gave the formula for the $\beta_{Y1.2}$ and $\beta_{Y2.1}$ coefficients as follows:

$$\beta_{Y1.2} = \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2}$$

$$\beta_{Y2.1} = \frac{r_{Y2} - r_{Y1}r_{12}}{1 - r_{12}^2} \qquad (5)$$

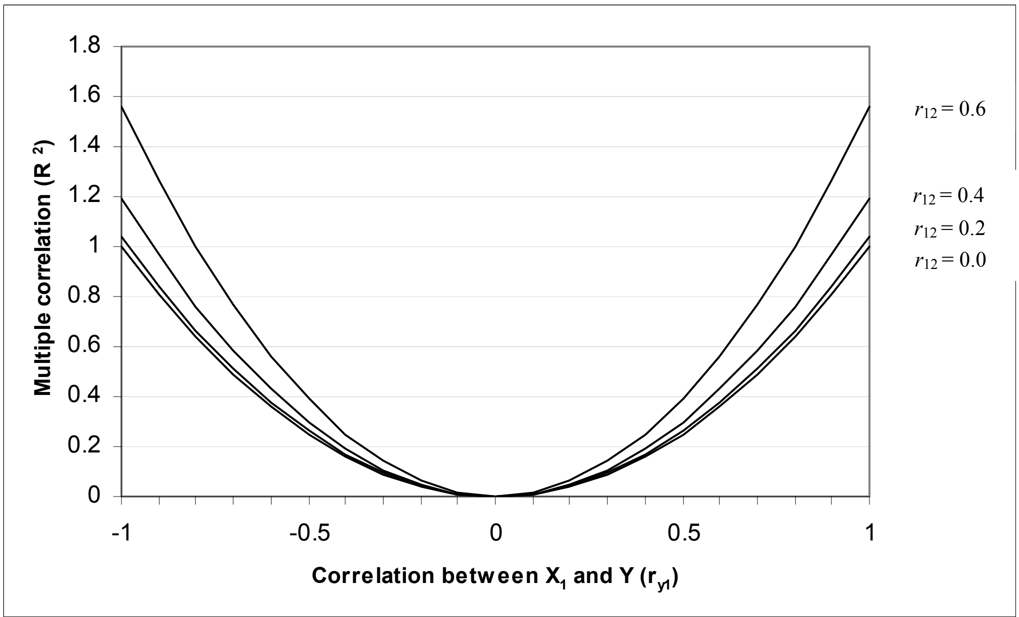

Figure 2. The inflation of $R^2_{Y.12}$

Since $r_{Y2} = 0$, Equation (5) can be simplified as

$$\beta_{Y1.2} = \frac{r_{Y1}}{1 - r_{12}^2} \quad \text{and} \quad \beta_{Y2.1} = \frac{-r_{Y1}r_{12}}{1 - r_{12}^2} \qquad (6)$$

The sign of $\beta_{Y2.1}$ depends on the sign of r_{12}. If there is a negative correlation between X_1 and X_2, the sign of $\beta_{Y2.1}$ will be the same as the sign of $\beta_{Y1.2}$. If there is a positive correlation between X_1 and X_2, the sign of $\beta_{Y2.1}$ and $\beta_{Y1.2}$ will be the opposite as can be seen in Figure 3. When $\beta_{Y2.1}$ has a positive sign, Krus and Wilkinson (1986) labelled it as 'positive classical suppression', and when $\beta_{Y2.1}$ has a negative sign they labelled it as 'negative classical suppression'. The magnitude of the inflations of $\beta_{Y2.1}$ and $\beta_{Y1.2}$ from their bivariate correlation with the criterion, r_{Y1} and r_{Y2} also depend on the value of r_{12}. A higher the value of

r_{12} leads to bigger inflations of $\beta_{Y1.2}$ and $\beta_{Y2.1}$ and beyond a certain point the value of $\beta_{Y1.2}$ and $\beta_{Y2.1}$ can exceed 1.

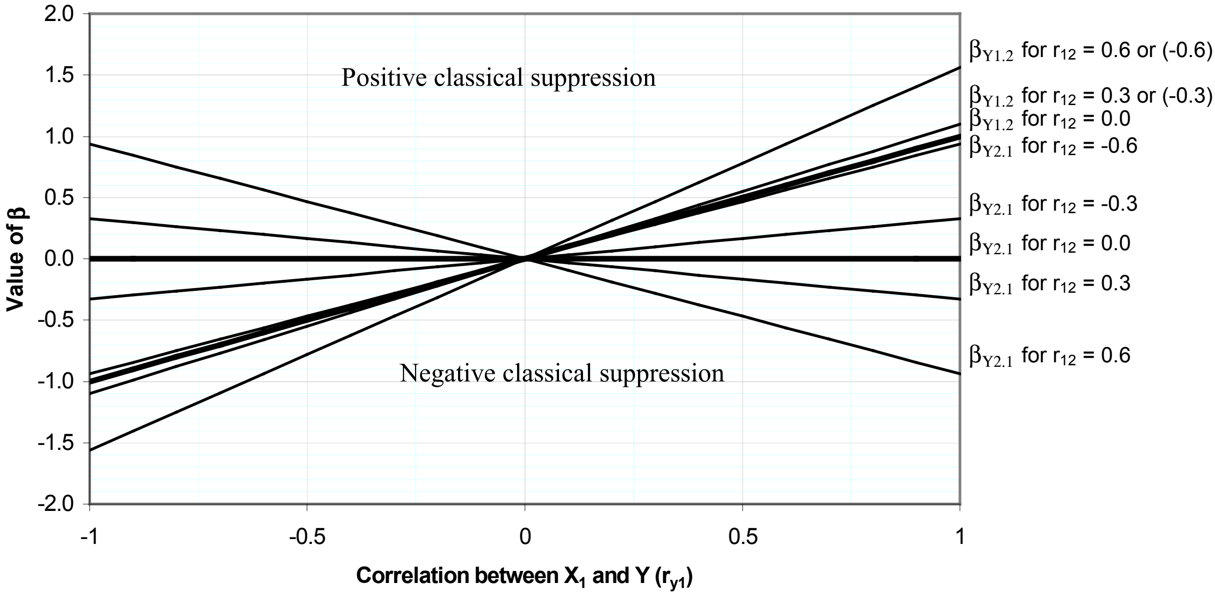

Figure 3. Classical suppression

Net suppression

This type of suppression occurs when a predictor variable has a regression weight with an opposite sign to its correlation with the criterion. In other word, $r_{Y1} \neq 0$, $r_{Y2} \neq 0$, and $r_{12} \neq 0$ but the $\beta_{Y2.1}$ is opposite in sign to r_{Y2}. In order to understand the meaning of these coefficients it is useful to consider the Venn diagram shown in Figure 4.

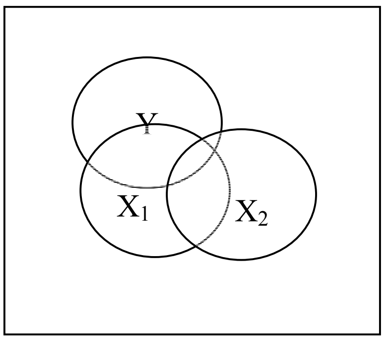

Figure 4. A Venn diagram for net suppression

Here the primary function of X_2 is to suppress the error variance X_1, rather than influencing substantially Y. As can be seen in Figure 4 X_2 has much more in common with the error variance in X_1 than it does with the variance in Y. This can happens when X_2 is highly correlated with X_1 but weakly correlated with Y.

In Figure 5 various $\beta_{Y2.1}$ values for $r_{12} = 0.6$ and $r_{12} = -0.6$ have been plotted. If X_2 is positively correlated with Y but has a negative value of $\beta_{Y2.1}$, Krus and Wilkinson (1986) labelled it as 'negative net suppression'. If X_2 is negatively correlated with Y but has a positive value of $\beta_{Y2.1,}$ Krus and Wilkinson (1986) called it 'positive net suppression'.

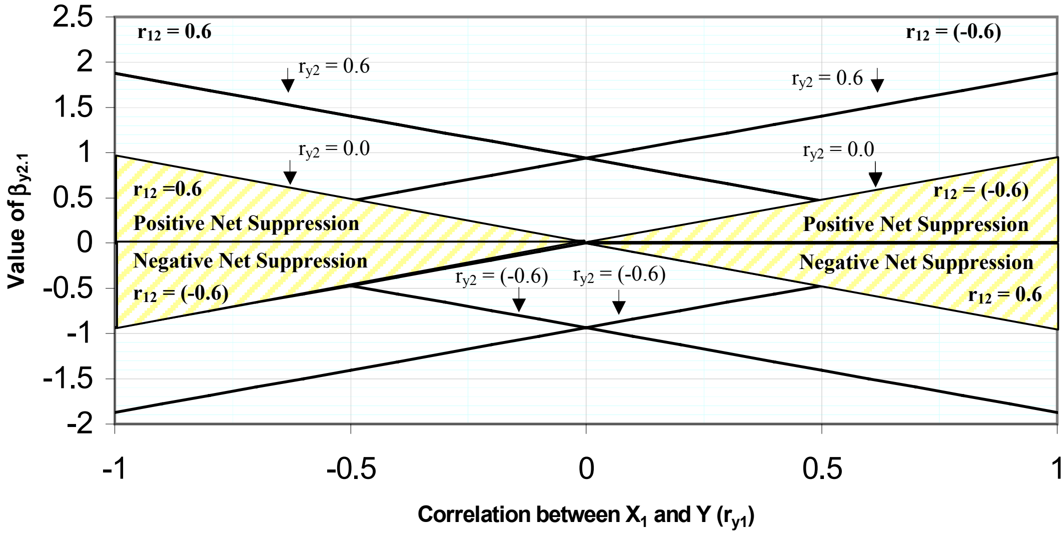

Figure 5. Net Suppression

Cooperative suppression

Co-operative suppression occurs when the two predictors are negatively correlated with each other, but both are positively or negatively correlated with *Y*. This is a case where each variable accounts for more of the variance in *Y* when it is in an equation with the other than it does when it is presented alone. As can be seen in Figure 6, when r_{12} is set to -0.6, the value of R^2 is more highly boosted as r_{Y2} increases. When both X_1 and X_2 are positively correlated with *Y*, Krus and Wilkinson (1986) labelled it as “positive cooperative suppression”; and when both X_1 and X_2 are negatively correlated with *Y*, Krus and Wilkinson (1986) labelled it as ’negative cooperative suppression’ as shown in Figure 7.

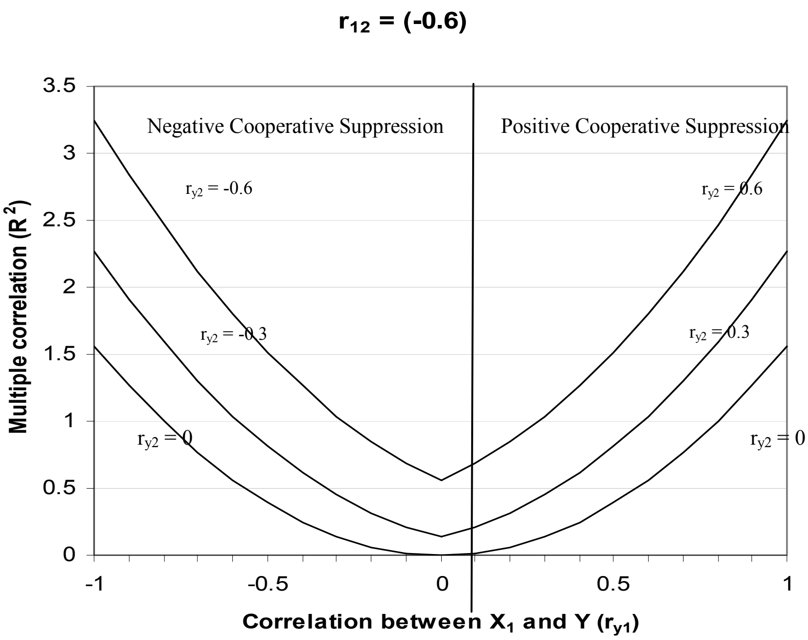

Figure 6. R^2 values in Cooperative Suppression

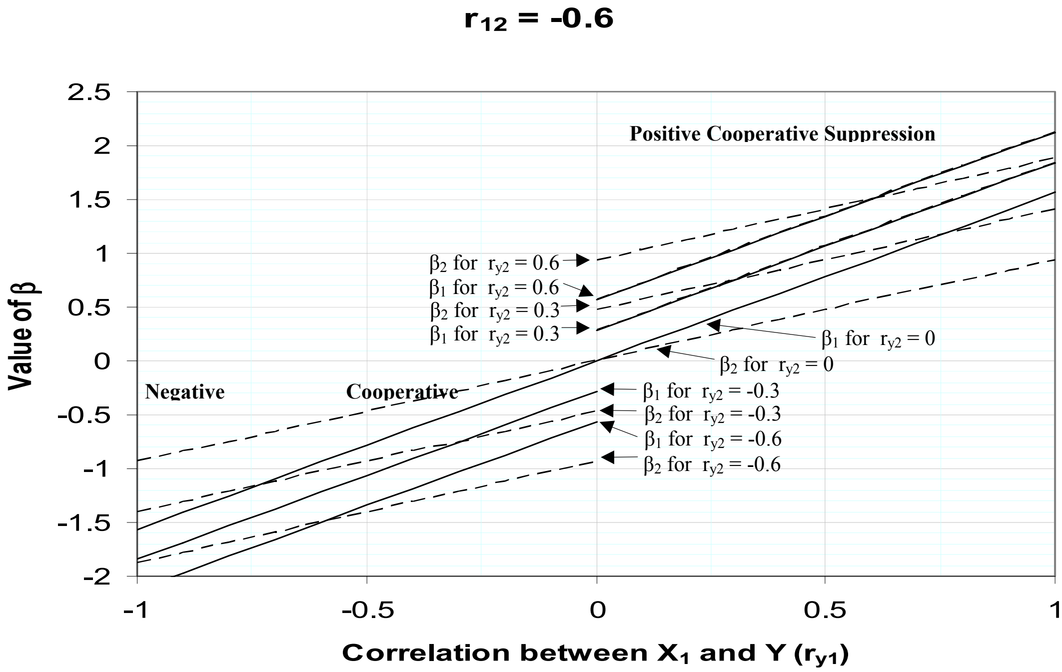

Figure 7. Cooperative Suppression

Cohen and Cohen (1983) suggested that one indication of suppression is a standardised regression coefficient (β_i) that falls outside the interval $0 < \beta_i < r_{Yi}$. To paraphrase Cohen and Cohen (1983), if X_i has a (near) zero correlation with Y, then there is possible classical suppression present. If its b_i is opposite in sign to its correlation with Y, there is net suppression present. And if its b_i exceeds r_{Yi} and it has the same sign, there is cooperative suppression present.

Multicollinearity has adverse effects not only on the regression and the multiple correlation coefficients, but also on the standard errors of regression coefficients as well as on the accuracy of computations due to rounding errors. In order to detect such problems concepts of a 'variance inflation factor' (VIF) and 'tolerance' were introduced (Pedhazur, 1997; Cohen et al., 2003).

$$VIF_i = \frac{1}{1-R_i^2}$$
$$Tolerance = \frac{1}{VIF_i} = 1 - R_i^2 \tag{7}$$

For a regression with two independent variables:

$$R_1^2 = R_2^2 = 1 - \frac{1}{\left(\frac{1}{1-r_{12}^2}\right)} = 1-(1-r_{12}^2) = r_{12}^2 \tag{8}$$

$$VIF_1 = VIF_2 = \frac{1}{1-r_{12}^2}$$
$$Tolerance_1 = Tolerance_2 = 1 - r_{12}^2 \tag{9}$$

The smaller the tolerance or the higher the VIF, the greater are the problems arising from multicollinearity. There is no agreement on cut-off values of tolerance. BMDP uses a tolerance of 0.01 as a default cut-off for entering variables, MINITAB and SPSS use a default value of 0.0001 (Pedhazur, 1997, p. 299). Cohen et al. (2003, p. 423) suggested that any VIF of 10 or more provides evidence of serious multicollinearity, which is equal to a tolerance of 0.1. Furthermore, they argued that "the values of the multicollinearity indices at which the interpretation of regression coefficients may become problematic will often be considerably smaller than traditional rule of thumb guidelines such as VIF =10". Sellin (1990) used the squared multiple correlation between a predictor and the set of remaining predictors involved in the equation (R_i^2) to indicate the relative amount of multicollinearity, He mentioned that relatively large values, typically those larger than 0.5, which is equal to VIF = 2, may cause problems in the estimation.

SOME ALTERNATIVE STRATEGIES

When a researcher is concerned only with the prediction of Y, multicollinearity has little effect and no remedial action is needed (Cohen et al., 2003 p.425). However, if interest lies in the value of regression coefficients or in the notion of causation, multicollinearity may introduce a potentially serious problem. Pedhazur (1997) and Cohen et al. (2003) proposed some strategies to overcome this problem that included (a) model respecification, (b) collection of additional data, (c) using ridge regression, and (d) principal components regressions.

When two or more observed variables are highly correlated, it may be possible to create a latent variable, that can be used to represent a theoretical construct which cannot be observed directly. The latent construct is presumed to underlie those observed highly correlated variables (Byrne, 1994).

The authors of this article have focused on this strategy, to create latent constructs and to extend the hierarchical linear model to accommodate the latent constructs. It also seeks to include indirect paths into the hierarchical linear model with the latent predictor. Thus, an attempt has been made to combine the strengths of the two common approaches in analysing multilevel data: (a) two-level path models that can estimate direct and indirect effects at two levels, can use latent constructs as predictor variables, but can not estimate any cross-level interaction; and (b) hierarchical linear models that can estimate direct and cross-level interaction effects, but can not estimate indirect paths nor use latent constructs as predictor variables. Muthén and Muthén (2004) have developed a routine called 'multilevel mixture modelling' that can estimate a two-level model which has latent constructs as predictor variables, direct and indirect paths, as well as cross-level interactions.

DATA AND VARIABLES

The data used in this study were collected from 1,984 junior secondary students in 71 classes in 15 schools in Canberra, Australia. Information was collected about individual student socioeconomic status (father's occupation), student aspirations (expected occupation, educational aspirations (expected education), academic motivation, attitude towards science (like science), attitude towards school in general (like school), self-regard, prior science achievement and final science achievement (outcome). In addition, information on class sizes was also collected. The outcome measure was the scores on a science achievement test of 55 items.

The names, codes and description of the predictor variables tested for inclusion at each level have been given in Table 1.

Table 1: Variables tested at each level of the hierarchy

Level	Variable code	Variable description
Level-1		(Student-level)
Student Background (N=1984)	FOCC	Father's occupation (1=Professional, . . . , 6=Unskilled labourer)
	EXPOCC	Expected occupation (1=Professional, . . . , 6=Unskilled labourer)
	EXPED	Expected education (1=Year 10 and Below, . . . ; 6=Higher Degree)
	ACAMOT	Academic motivation (0=Lowest motivation, . . . , 40=Highest motivation)
	LIKSCH	Like school (0=Likes school least, . . . , 34=Likes school most)
	LIKSCI	Like science (1=Likes science least, . . . , 40=Likes science most)
	SELREG	Self regard (1=Lowest self regard, . . . , 34=Highest self regard)
	ACH68	Prior science achievement (0=Lowest score, . . . , 25=Highest score)
Level-2		(Class-level)
Class Characteristics Group Composition (n=71)	CSIZE	Class size (8=Smallest, . . . , 39=Largest)
	FOCC_2	Average father occupation at class-level
	EXPOCC_2	Average expected occupation at class-level
	EXPED_2	Average expected education at class-level
	ACAMOT_2	Average academic motivation at class-level
	LIKSCH_2	Average like school at class-level
	LIKSCI_2	Average like science at class-level
	SELREG_2	Average self regard at class-level
	ACH68_2	Average prior science achievement
Outcome	ACH69	Science Achievement (1 =lowest score….55=highest score)

HLM MODEL: THE INITIAL MODEL

Initially a two-level model was fitted using HLM 6. The first step in the HLM analyses was to run a fully unconditional model in order to obtain the amounts of variance available to be explained at each level of the hierarchy (Bryk and Raudenbush, 1992). The fully unconditional model contained only the dependent variable (Science achievement, ACH) and no predictor variables were specified at the class level. The fully unconditional model is stated in equation form as follows.

Level-1 model

$$Y_{ij} = \beta_{0j} + e_{ij}$$

Level-2 model

$$\beta_{0j} = \gamma_{0j} + r_{0j} \qquad (10)$$

where:

Y_{ij} is the science achievement of student *i* in class *j*;

The second step undertaken was to estimate a Level-1 model, that is, a model with student-level variables as the only predictors in Equation 10. This involved building up the student-level model or the so-called 'unconditional' model at Level-1 by adding student-level predictors to the model, but without entering predictors at the other level of the hierarchy. At this stage, a step-up approach was followed to examine which of the eight student-level variables (listed in Table 1) had a significant (at p≤0.05) influence on the outcome variable, ACH69. Four variables (FOCC, EXPED, LIKSCI and ACH68) were found to be significant and therefore were included in the model at this stage. These four student-level variables were grand-mean-centred in the HLM analyses so that the intercept term would represent the ACH69 score for student with average characteristics.

The final step undertaken was to estimate a Level-2 model, which involved adding the Level-2 or class-level predictors into the model using the step-up strategy mentioned above. At this stage, the Level-2 exploratory analysis sub-routine available in HLM 6 was employed for examining the potentially significant Level-2 predictors in successive HLM runs. Following the step-up procedure, two class-level variables (CSIZE and ACH68_2) were included in the model for the intercept. In addition, one cross-level interaction effect between ACH68 and CSIZE was included in the model.

The final model at Levels 1, and 2 can be denoted as follows.

Level-1 Model

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(\text{FOCC}) + \beta_{2j}*(\text{EXPED}) + \beta_{3j}*(\text{LIKSCI}) + \beta_{4j}*(\text{ACH68}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\text{ACH68\_2}) + \gamma_{02}*(\text{CSIZE}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}*(\text{CSIZE}) + u_{4j} \quad (11)$$

The next step was to re-estimate the final model using the MPLUS program. The results of the estimates of fixed effects from the two-level model are given in Table 2 for HLM and MPLUS estimation.

RESULTS

At the student-level, from the results in Table 2 it can be seen that Science achievement was directly influenced by Father's occupation (FOCC), Expected education (EXPED), Like science (LIKSCI) and Prior achievement (ACH68). When other factors were equal, students whose fathers had high status occupations (e.g. medical doctors and lawyers) outperformed students whose fathers had low status occupations (e.g. labourer and cleaners). Students who aspired to pursue education to high levels were estimated to achieve better when compared to students who had no such ambitions, while students who liked science were estimated to achieve better when compared to students who did not like science. In addition, students who had high prior achievement scores were estimated to achieve better than students who had low prior achievement scores.

At the class-level, from the results in Table 2 it can be seen that Science achievement was directly influenced by Average prior achievement (ACH68_2) and Class size (CSIZE). When other factors were equal, students in classes with high prior achievement scores were likely to achieve better when compared to students in classes with low prior achievement scores. Importantly, there was considerable advantage (in term of better achievement in science) associated with being in larger classes. These relationships have been shown in Figure 8.

From the results in Table 2 it can also be seen that there is one significant cross-level interaction effect ACH68 and CSIZE. This interaction is presented in Figure 9. Nevertheless, in interpreting the effects of class size, it should be noted that 10 out of the 15 schools in these data had a streaming policy that involved placing high achieving students in larger classes and low achieving students in smaller classes for effective teaching. Therefore, the better performance of the students in larger classes in these data was not surprising.

Table 2. HLM and MPLUS results for initial model

Level 1 N=1984	Level 2 n=71	HLM Estimate (se)	MPLUS Estimate (se)
Intercept		28.37 (0.20)	28.87 (0.19)
	ACH68_2	0.78 (0.10)	0.76 (0.12)
	CSIZE	0.16 (0.04)	0.16 (0.04)
FOCC		-0.25 (0.09)	-0.24 (0.10)
EXPED		0.48 (0.09)	0.49 (0.09)
LIKSCI		0.15 (0.01)	0.15 (0.01)
ACH68		0.91 (0.04)	0.93 (0.04)
	CSIZE	0.013 (0.005)	0.015 (0.006)

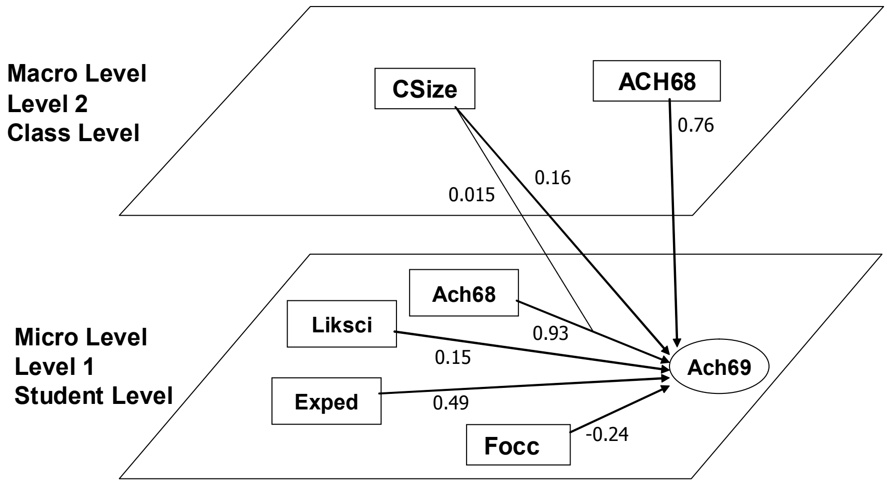

Figure 8. Model 1: Initial Model (MPlus results used)

ALTERNATIVE MODELS

Two alternative models, Model 2 and Model 3, were estimated using MPLUS 3.13. Both EXPED and EXPOCC are significantly correlated with ACH69 with correlation coefficients of 0.50 and 0.35 respectively. Either EXPED or EXOCC can have a significant effect on ACH69. However, if the two variables were put together as predictors of ACH69, only EXPED was found to be significant. Since there is a relatively high correlation between EXPOCC and EXPED (-0.53) it is possible to form a latent construct, labelled as aspiration (ASP), and use this construct as a predictor variable instead of just using either EXPOCC or EXPED. In this way, both variables (EXPOCC and EXPED) become significant reflectors of aspiration. Otherwise, EXPOCC may be regarded as an insignificant predictor of science achievement as in the initial model. The results have been recorded in Table 3 and Model 2 is shown visually in Figure 10. This employment of a latent construct is very useful in situations where three observed predictor variables are available and suppressor relationships occur if all three predictor variables are introduced separately into the regression equation.

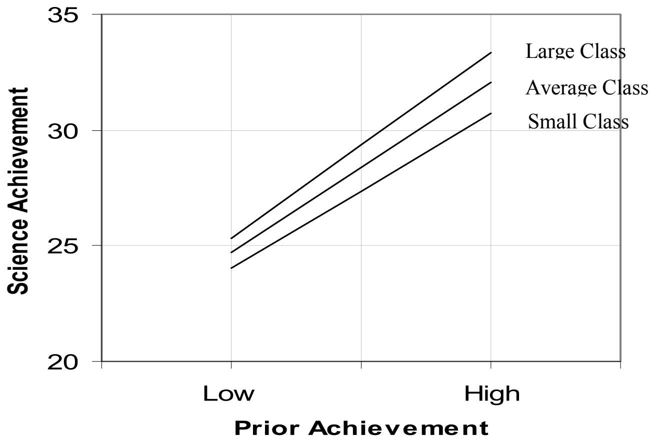

Figure 9. Interaction effect between CSIZE and PRIORACH

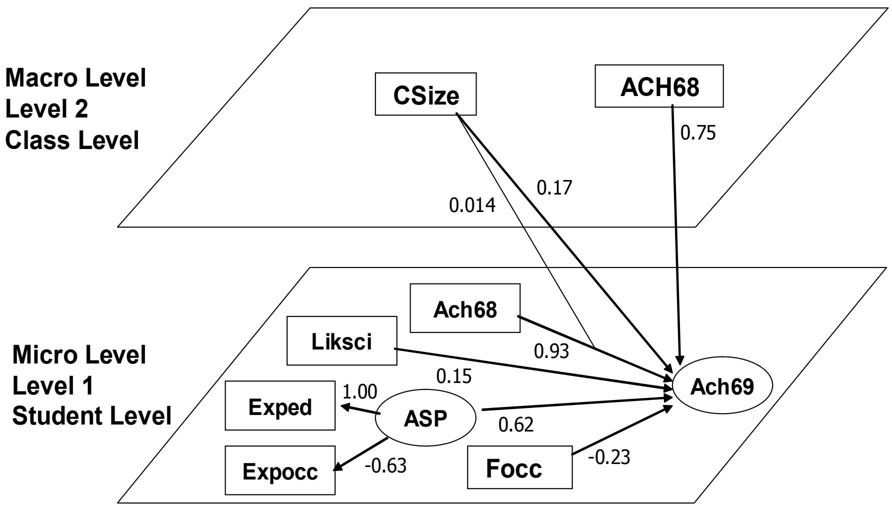

Figure 10. Model 2: With latent construct

The next step undertaken was to estimate another model with two additional indirect paths. It was hypothesised that academic motivation (ACAMOT) influenced like science at the student level and average father's occupational status influences average prior achievement at the class level. The results are recorded in Table 3 and Model 3 is shown in Figure 11.

The proportions of variance explained at each level for each model are presented in Table 4. For Model 1, the initial model, 45 per cent of variance available at Level 1 and almost all (95 %) of variance available at Level 2 have been explained by the inclusion of four variables at Level 1 (FOCC, EXPED, LIKSCI, and ACH68) and two variables at Level 2 (ACH68 and CSIZE) as well as one interaction effect between ACH68 and CSIZE. Overall this model explained 68.7 per cent of total variance available when the model was estimated with HLM. MPLUS estimations are very close to HLM estimations. Adding a latent construct into the model did not really increase the amount of variance explained, but it did give a more coherent picture of the relationships. This is also true for Model 3 when indirect paths are added.

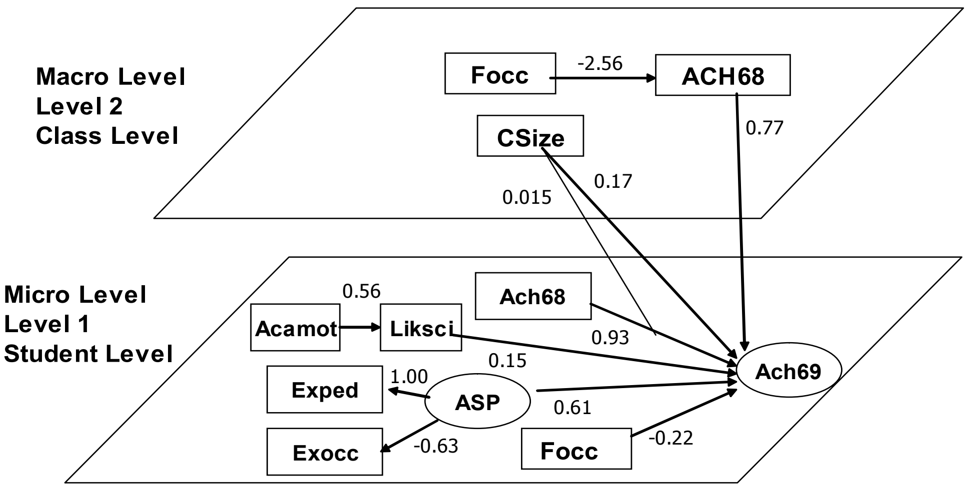

Figure 11. Model 3: With latent construct and indirect paths

Table 3. Model 2 and Model 3 Results

Level 1 (N=1984)	**Level 2 (n=71)**	**Model 2 with latent construct**	**Model 3 with latent construct and indirect paths**
Criterion ACH69		**estimate (se)**	**estimate (se)**
Latent Construct			
ASP by			
EXPED		1.00(0.00)	1.00(0.00)
EXPOCC		-0.63(0.10)	-0.63(0.11)
Indirect Paths			
ACAMOT on LIKSCI			0.56 (0.03)
	FOCC_2 on ACH68		-2.56 (0.30)
Fixed Effects			
Intercept		28.87 (0.20)	28.85 (0.20)
	ACH68	0.75 (0.12)	0.77 (0.12)
	CSIZE	0.17 (0.05)	0.17 (0.05)
FOCC		-0.23 (0.10)	-0.22 (0.10)
ASP		0.62 (0.14)	0.61 (0.14)
LIKSCI		0.15 (0.01)	0.15 (0.01)
ACH68		0.93 (0.04)	0.93 (0.04)
	CSIZE	0.014 (0.007)	0.015 (0.01)

CONCLUSIONS

Multicollinearity is one of the problems that need to be examined carefully when a multiple regression model is employed. When the main concern is merely the prediction of *Y*, multicollinearity generally has little effect, but if the main interest lies in the value of regression coefficients, multicollinearity may introduce a potentially serious problem.

Multilevel mixture modelling, which can estimate a two-level model that has latent constructs as predictor variables, direct and indirect paths, as well as cross-level interactions, has been used as an alternative strategy to analyse multilevel data. In a sense, this approach can be seen as an attempt to combine the strengths of the two commonly used techniques in analysing multilevel data, two level path modelling and hierarchical linear modelling.

The initial model was a hierarchical linear model, which was fitted using both HLM 6 and MPLUS 3.13. Both estimations yielded similar results. The main effects reported from the analysis at the student-level, indicate that in addition to prior achievement, it was the social psychological measures associated with the differences between students within classrooms that were having effects, namely, socioeconomic status, educational aspirations, and attitudes towards learning science. About 55 per cent of the variance between students within classrooms was left unexplained, indicating that there were other student-level factors likely to be involved in influencing student achievement.

Table 4. Variance components

	HLM			MPLUS		
Model (N=1984, n=71)	**Level 1**	**Level 2**	**Total**	**Level 1**	**Level 2**	**Total**
Null Model						
Variance Available	38.07	33.85	71.92	38.07	33.35	71.42
Initial Achievement (Residual)	24.25	9.34	33.59	24.33	8.45	32.78
Total Variance Explained %	36.3	72.4	53.3	36.1	74.7	54.1
Total Variance Unexplained %	63.7	27.6	46.7	63.0	25.3	45.9
Model 1: Initial Model (Residual)	20.93	1.60	22.53	21.01	1.46	22.46
Total Variance Explained %	45.0	95.3	68.7	44.8	95.6	68.6
Total Variance Unexplained %	55.0	4.7	31.3	55.2	4.4	31.4
Model 2: With Latent Predictor (Residual)				21.36	1.49	22.84
Total Variance Explained %				43.9	95.5	68.0
Total Variance Unexplained %				56.1	4.5	32.0
Model 3: Add indirect Paths (Residual)				21.36	1.50	22.85
Total Variance Explained %				43.9	95.5	68.0
Total Variance Unexplained %				56.1	4.5	32.0

At the classroom level, about 4.7 per cent of the variance between classes was left unexplained, with the average level of prior achievement of the class group had a significant effect. In addition, class size had a positive effect on science achievement, with students in larger classes doing significantly better than students in smaller classes. Perhaps, this indicates the confounding effect of streaming policy adopted by some schools to place better students in larger classes. In addition, the interaction effect also reveals that the effect of prior achievement is stronger in larger classes. High achieving students are better off in larger classes.

The next step was to add a latent construct, aspiration to the initial model. The estimation of this model was done by using the two-level mixture model procedure in MPLUS 3.13. By creating this latent construct, it could be said that aspiration, which was reflected significantly by expected education and expected occupation, had a positive effect on achievement.

The last step was to add two indirect paths, one at the student level and one at the class level. At the student level, academic motivation was found to have a significant effect on like science and indirectly influence achievement through like science. At the class level, average fathers' occupation was related to average prior achievement.

By using multilevel mixture modelling, the limitations of hierarchical linear modelling are partly reduced. The ability to include latent constructs in a path model reduces the problem of multicollinearity and multiple measures. The inclusion of indirect paths also increases the modelling possibilities. However, these estimations need greater computing power if larger models are to be examined.

REFERENCE

Bryk, A.S. and Raudenbush, S.W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA.: Sage Publications.

Byrne, B.M. (1994) *Structural Equation Modelling with EQS and EQS/Windows: Basic Concepts, Applications, and Programming*, Thousand Oaks, CA: Sage Publications.

Cohen, J. and Cohen, P. (1975). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. New York: Wiley.

Cohen, J. and Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences,* (2^{nd} ed.). Hillsdale, NJ. : Erlbaum Associates.

Cohen, J., West, S.G., Aiken, L. and Cohen, P. (2003) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3^{rd} ed. Mahwah, NJ.: Erlbaum Associates

Conger, A.J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, 34, 35-46.

Gustafsson, J.E., and Stahl, P.A. (1999). *STREAMS User's Guide Version 2.5 for Windows*, Molndal, Sweden: Multivariate Ware.

Horst, P. (1941). The role of predictor variables which are independent of the criterion. *Social Science Research Bulletin*, 48, 431 -436.

Hox, J.J. (1994). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Krus, D.J. and Wilkinson, S.M. (1986). Demonstration of properties of a suppressor variable. *Behavior Research Methods, Instruments, and Computers*, 18, 21-24.

Muthén, L.K. and Muthén, B.O. (2004). *Mplus: The Comprehensive Modelling Program for Applied Researchers. User's guide* (3rd ed.). Los Angeles: Muthén and Muthén.

Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research: Explanation and Prediction* (3rd ed.). Forth Worth, TX.: Harcourt Brace College Publishers.

Sellin, N. (1990). *PLSPATH Version 3. 01. Application Manual*. Hamburg, Germany.

❖IEJ