International Education Journal, 2006, 7(2), 174-188.
ISSN 1443-1475 © 2006 *Shannon Research Press*.
http://iej.cjb.net

Accountability of teachers and schools: A value-added approach

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*

Currently, there has been substantial interest, in Australia and internationally, in policy activities related to outcomes-based educational performance indicators and their link with growing demands for accountability of teachers and schools. In order to achieve a fair comparison between schools, it is commonly agreed that a correction should be made for lack of equity. It is argued that student performance is influenced by three general factors: the student background, classroom and school context, and identified school policies and practices. In this article the effects of these three factors on science achievement among students in Canberra, Australia have been addressed. The effects are discussed with reference to Type A, Type B, Type X, and Type Z effects. Type A effects are school effectiveness indicators controlled for student background. Type B school effects are controlled for both student background and context variables. Type X effects are estimated with student effects, context effects and non-malleable policy effects controlled for. Finally, Type Z effects invoke school effectiveness indicators, controlled for student, context, and all identified policy effects.

Value-added, accountability, science achievement, social psychological measures, equity, school effectiveness indicators

ACCOUNTABILITY OF TEACHERS AND SCHOOLS

During the past two decades there has been a growing interest in the performance and accountability of teachers and schools both in Australia and internationally (Rowe, 2000). Educational outcome indicators are frequently used to measure the performance of teachers, schools, programs, and policies. Reliance on such indicators is largely the result of a growing demand to hold these entities accountable for performance, defined in terms of outcomes, such as standardised test scores in science, rather than inputs such as student prior achievement, teacher quality, class size, or quality of facilities (Meyer, 2000, 2002). The use of such indicators, for example average or median test scores, has some major shortcomings. Rowe (2000) pointed out that the analyses of test scores tended to be focused on a comparative ranking of schools rather than on identifying factors that explained school differences. Moreover, Meyer (2002) contended that average test scores (a) were influenced by factors other than school performance; (b) were a reflection of the accumulated learning that had occurred; (c) tended to be contaminated due to student mobility; and (d) failed to localise school performance to a specific classroom or grade level.

Given these problems associated with the use of common educational outcome indicators, the papers by Ballou et al. (2004), De Fraine et al. (2002), Raudenbush and Willms (1995), Rubin et al. (2004), and Willms and Raudenbush (1989) have approached the estimation of school and teacher effects through the use of a variety of statistical models, known as 'value-added' models in the education literature. The essence of the value-added approach is to isolate statistically the contribution of teachers and schools to growth in student achievement at a given grade level from all other sources of student achievement growth. Failure to isolate these contributions could result in highly contaminated indicators of performance.

Consequently, the emphasis in cross-national achievement surveys, as well as national studies of educational achievement that compare the performance of schools using the rank ordering or scaling of outcomes fail to examine in a meaningful way differences in performance unless further analyses that estimate value-added effects are carried out.

FOUR TYPES OF SCHOOL EFFECTS

Type A, Type B, Type X, and Type Z Effects

Raudenbush and Willms (1995, p. 313) and Willms and Raudenbush (1989, pp. 212-214) argued that student performance (Y) was influenced by three general factors: the student background characteristics (S), school context (C) and identified school policies, practices, and stratifications (P), as well as each student's unique contribution (e).

$$Y_{ij} = \mu_{0j} + S_{ij} + C_{ij} + P_{ij} + e_{ij} \qquad (1)$$

This model can be extended to accommodate classroom or teacher effects by splitting school context (C) into its components, namely classroom context (CC) and school context (SC). Furthermore, school policies and practices (P) can be divided into identified policies and practices (IP) and unidentified policies and practices (UP). Identified policies and practices (IP) can be further subdivided into malleable policies and practices (MP) and non-malleable policies and practices (NP). It should be noted that non-malleable polices and practices (NP) can be identified, but a school has no control over them since they are determined at the system level, while malleable policies and practices (MP) are under a school's control. Hence we may write

$$Y_{ijk} = \mu_{0jk} + S_{ijk} + CC_{ijk} + SC_{ijk} + NP_{ijk} + MP_{ijk} + UP_{ijk} + e_{ijk} \qquad (2)$$

Equation (2) can also be written with further error terms (u_{00k} and r_{0jk}) included:

$$Y_{ijk} = \gamma_{000} + S_{ijk} + CC_{ijk} + SC_{ijk} + NP_{ijk} + MP_{ijk} + UP_{ijk} + u_{00k} + r_{0jk} + e_{ijk} \qquad (3)$$

Four types of teachers or school effects can be distinguished: Type A, Type B effects (Raudenbush and Willms, 1995; Willms and Raudenbush, 1989), Type X effects (Hungi, 2003; Keeves et al., 2005) and Type Z effects.

Type A effects refer to how well the students in a school perform in comparison with the performance of similar students in other schools. Type A effects are of interest for students and parents in choosing a school. Parents want to know which school can help their child to excel. Parents and students will choose the school with the largest Type A effect, that is the school with the largest value added effect when individual student characteristics are taken into account. The Type A effects can be specified as:

$$A_{ijk} = CC_{ijk} + SC_{ijk} + NP_{ijk} + MP_{ijk} + UP_{ijk} + u_{00k} + r_{0jk} + e_{ijk} \qquad (4)$$

Type B effects refer to how well the students in a classroom within a school perform, compared to similar students in classrooms and schools with similar contexts. Type B effects are of interest for those who are looking for accountability of the teacher and school. Teachers and principals are more interested in the Type B effects of their own schools because they look for an indication of their school's performance, excluding factors that lie beyond their control. Type B effects are also of interest for administrators and education policy makers, looking for accountability. Schools should not be held accountable for the context in which they operate. The Type B effects can be specified as:

$$B_{ijk} = NP_{ijk} + MP_{ijk} + UP_{ijk} + u_{00k} + r_{0jk} + e_{ijk} \tag{5}$$

There are some non-malleable polices and practices (NP), polices and practices that can be identified but the school has no control over them, and they should be removed, such as whether the school is urban or rural, or the size of the school in situations where the school has no control over its size, as well as other stratifying variables such as State or School Type. Therefore, Type X effects refer to how well the students in a classroom within a school perform, when compared to similar students in classrooms and schools with similar contexts as well as similar non-malleable policies and practices. It may be argued that the Type X estimate is the most appropriate estimate of value added, with student effects, context effects (CC and CS), and identified non-malleable policy effects (NP) removed from the value added estimates. Type X effects can be specified as:

$$X_{ijk} = MP_{ijk} + UP_{ijk} + u_{00k} + r_{0jk} + e_{ijk} \tag{6}$$

However, it would seem appropriate to judge a school by the effect of identified malleable polices and practices as well. An example of malleable policy and practice at the school level would seem to be that of 'streaming'. After controlling for the malleable policy and practices, the remaining effects can be labelled as Type Z effects and can be written as

$$Z_{ijk} = UP_{ijk} + u_{00k} + r_{0jk} + e_{ijk} \tag{7}$$

DATA SAMPLE

The data used in this study were collected from 1,984 junior secondary students in 71 classes in 15 schools in Canberra, Australia. These 15 schools consisted of nine government schools, four Catholic schools and two independent schools. Nine of these schools were co-educational schools and six were single sex (three boys' and three girls' schools). In addition, ten out of the 15 schools had a streaming policy of placing high achieving students in larger classes. The sample represents a cohort of approximately 2000 students, who transferred from Grade 6 to Grade 7 within a small school system.

HYPOTHESIZED MODEL

Testing of hypotheses in multilevel models can be carried out using multilevel data analyses software such as HLM 6 for Windows (Raudenbush et al., 2004). The HLM program was initially developed to find a solution for the methodological weakness of educational research studies during the early 1980s, which was the failure of many analytical studies to attend to the hierarchical, multilevel character of much of educational field research data (Bryk and Raudenbush, 1992). This failure came from the fact that "the traditional linear models used by most researchers require the assumption that subjects respond independently to educational programs" (Raudenbush and Bryk; 1994, p. 2590). In practice, most educational research studies select students as a sample who are nested within classrooms, and the classrooms are in turn nested within schools, and schools exist within geographical regions. In this situation, the students selected in the study are not independent, but rather nested within organisational units and ignoring this fact results in the problems of "aggregation bias and misestimated precision" (Raudenbush and Bryk, 1994, p. 2590).

In Figure 1 the three-level model proposed for testing in this study is shown. The names, codes and description of the predictor variables tested for inclusion at each level of the three-level model have been provided in Table 1. Apart from Class size (CSIZE) at class level and school classifications at school level, all the other variables at the class and school levels were constructed by aggregating the student-level data.

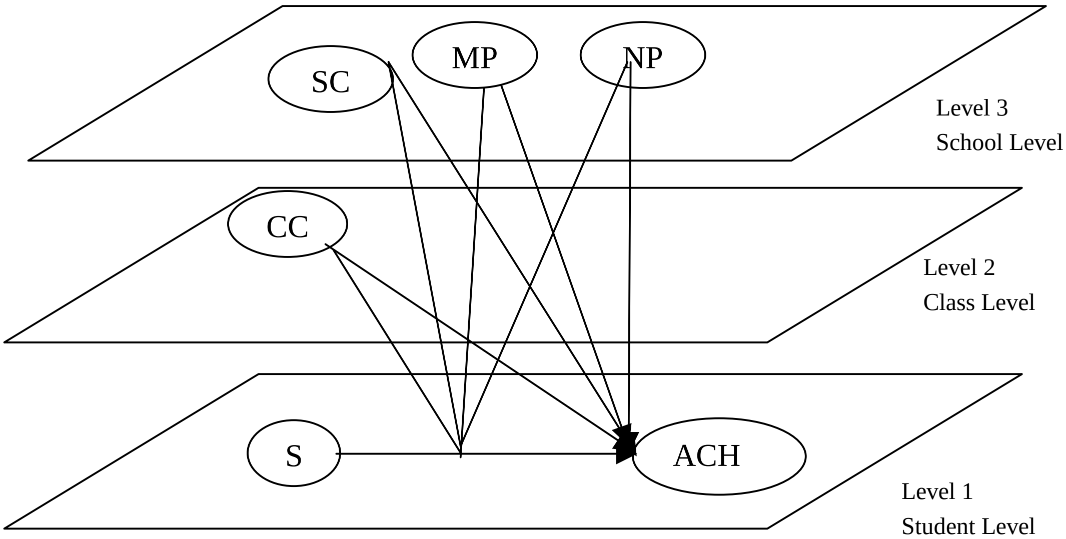

Figure 1. Hypothesised three-level hierarchical model for science achievement

ANALYSES

The multilevel models were built step-by step. The first step was to run a model without explanatory variables, which is also called the 'null model'. Thus null model was fitted to provide estimates of the variance components at each level (Raudenbush and Bryk , 2002). The null model can be stated in equation form as follows.

Level-1 model

$$Y_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 model

$$\pi_{0jk} = \beta_{00j} + r_{0jk}$$

Level-3 model

$$\beta_{00k} = \gamma_{000} + u_{00k} \qquad (8)$$

where: Y_{ikj} is the science achievement of student i in class j in school k.

The second step undertaken was to estimate Type A effects in which student characteristics were added, thereby controlling for student intake. 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 $p \leq 0.05$) influence on the outcome variable, ACH. Four variables (FOCC, EXPED, LIKSCI and PRIORACH) 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 average ACH score for the students with average student characteristics. When a variable was centred around its grand mean, the zero value indicated its average value.

Table 1. Variables tested at each level of the hierarchy

Level	Variable code	Variable description
Level-1		(Student-level)
(S)	FOCC	Father's occupation (1=Unskilled labourer,. . . , 6= Professional)
	EXPOCC	Expected occupation (1=Unskilled labourer,. . . , 6= Professional)
	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)
	PRIORACH	Prior science achievement (0=Lowest score, . . . , 25=Highest score)
Level-2		(Class-level)
(CC)	CSIZE	Class size (8=Smallest, . . . , 39=Largest)
	FOCC_2	Average fathers' 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
	PRIOR_2	Average prior science achievement
Level-3		(School-level)
(SC)	CSIZE_3	Average class size
	FOCC_3	Average fathers' occupation at school-level
	EXPOCC_3	Average expected occupation at school-level
	EXPED_3	Average expected education at school-level
	ACAMOT_3	Average academic motivation at school-level
	LIKSCH_3	Average like school at school-level
	LIKSCI_3	Average like science at school-level
	SELREG_3	Average self regard at school-level
	PRIOR_3	Average prior science achievement
(NP)	GOVT	Government school (0=Non-government; 1=Government)
	CATH	Catholic school (0=Non-Catholic; 1=Catholic)
	IND	Independent school (0=Non-Independent; 1=Independent)
	BOYS	Boys' school (0=Girls and Co-ed; 1=Boys only)
	GIRLS	Girls' school (0=Boys' and Co-ed; 1=Girls only)
	COED	Co-educational school (0=Boys only and Girls' only; 1=Co-ed)
(MP)	STREAM	Streaming in school (0=No streaming; 1=Streaming)
Outcome	ACH	Science Achievement (1 =lowest score….55=highest score)

The third step undertaken was to estimate Type B effects, which involved adding the classroom context and school context variables into the model using the step-up strategy mentioned above. At this stage, the Level-2 and Level-3 exploratory analysis sub-routines available in HLM 6 were employed for examining the potentially significant classroom and school context variables (as found in the output) in successive HLM runs. Following the step-up procedure, two classroom context variables (PRIOR_2 and CSIZE) were included in the model for the intercept. In addition, two cross-level interaction effects (between PRIORACH and FOCC_2 and between PRIORACH and LIKSCI_3) were included in the model.

The fourth step involved adding the significant non-malleable school policies and practices into the model using the Level-3 exploratory analysis sub-routine and the step-up strategy. At this stage, two cross-level interaction effects (between FOCC and GOV and between EXPED and

IND) were included in the model. In addition, the estimated coefficients for FOCC_2 were fixed at the school level because the reliability estimate of this coefficient was below 0.10.

The final step involved adding the malleable school policy into the model (STREAM). Estimates of fixed effects for Types A, B, X, and Z models have been given in Table 2.

The Type A model can be denoted as follows.

Level-1 model

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}FOCC_{ijk} + \pi_{2jk}EXPED_{ijk} + \pi_{3jk}LIKSCI_{ijk} + \pi_{4jk}PRIORACH_{ijk} + e_{ijk}$$

Level-2 model

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$
$$\pi_{2jk} = \beta_{20k} + r_{2jk}$$
$$\pi_{3jk} = \beta_{30k} + r_{3jk}$$
$$\pi_{4jk} = \beta_{40k} + r_{4jk}$$

Level-3 model

$$\beta_{00k} = \gamma_{000} + u_{00k}$$
$$\beta_{10k} = \gamma_{100} + u_{10k}$$
$$\beta_{20k} = \gamma_{200} + u_{20k}$$
$$\beta_{30k} = \gamma_{300} + u_{30k}$$
$$\beta_{40k} = \gamma_{400} + u_{40k} \tag{9}$$

The Type B model can be denoted as follows.

Level-1 model

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}FOCC_{ijk} + \pi_{2jk}EXPED_{ijk} + \pi_{3jk}LIKSCI_{ijk} + \pi_{4jk}PRIORACH_{ijk} + e_{ijk}$$

Level-2 model

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}PRIOR\_2_{0jk} + \beta_{02k}CSIZE_{0jk} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$
$$\pi_{2jk} = \beta_{20k} + r_{2jk}$$
$$\pi_{3jk} = \beta_{30k} + r_{3jk}$$
$$\pi_{4jk} = \beta_{40k} + \beta_{41k}FOCC\_2_{4jk} + r_{4jk}$$

Level-3 model

$$\beta_{00k} = \gamma_{000} + u_{00k}$$
$$\beta_{01k} = \gamma_{010} + u_{01k}$$
$$\beta_{02k} = \gamma_{020} + u_{02k}$$
$$\beta_{10k} = \gamma_{100} + u_{10k}$$
$$\beta_{20k} = \gamma_{200} + u_{20k}$$
$$\beta_{30k} = \gamma_{300} + u_{30k}$$
$$\beta_{40k} = \gamma_{400} + \gamma_{401}LIKSCI\_3_{40k} + u_{40k}$$
$$\beta_{41k} = \gamma_{410} + u_{41k} \tag{10}$$

The Type X model can be denoted as follows.

Level-1 model

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}FOCC_{ijk} + \pi_{2jk}EXPED_{ijk} + \pi_{3jk}LIKSCI_{ijk} + \pi_{4jk}PRIORACH_{ijk} + e_{ijk}$$

Level-2 model

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}PRIOR\_2_{0jk} + \beta_{02k}CSIZE_{0jk} + u_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$
$$\pi_{2jk} = \beta_{20k} + r_{2jk}$$
$$\pi_{3jk} = \beta_{30k} + r_{3jk}$$
$$\pi_{4jk} = \beta_{40k} + \beta_{41k}FOCC\_2_{4jk} + r_{4jk}$$

Level-3 model

$$\beta_{00k} = \gamma_{000} + u_{00k}$$
$$\beta_{01k} = \gamma_{010} + u_{01k}$$
$$\beta_{02k} = \gamma_{020} + u_{02k}$$
$$\beta_{10k} = \gamma_{100} + \gamma_{101}GOV_{10k} + u_{10k}$$
$$\beta_{20k} = \gamma_{200} + \gamma_{201}IND_{20k} + u_{20k}$$
$$\beta_{30k} = \gamma_{300} + u_{30k}$$
$$\beta_{40k} = \gamma_{400} + \gamma_{401}LIKSCI\_3_{40k} + u_{40k}$$
$$\beta_{41k} = \gamma_{410} \tag{11}$$

The Type Z model can be denoted as follows.

Level-1 model

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}FOCC_{ijk} + \pi_{2jk}EXPED_{ijk} + \pi_{3jk}LIKSCI_{ijk} + \pi_{4jk}PRIORACH_{ijk} + e_{ijk}$$

Level-2 model

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}PRIOR\_2_{0jk} + \beta_{02k}CSIZE_{0jk} + u_{0jk}$$
$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$
$$\pi_{2jk} = \beta_{20k} + r_{2jk}$$
$$\pi_{3jk} = \beta_{30k} + r_{3jk}$$
$$\pi_{4jk} = \beta_{40k} + \beta_{41k}FOCC\_2_{4jk} + r_{4jk}$$

Level-3 model

$$\beta_{00k} = \gamma_{000} + \gamma_{001}STREAM_{00k} + u_{00k}$$
$$\beta_{01k} = \gamma_{010} + u_{01k}$$
$$\beta_{02k} = \gamma_{020} + u_{02k}$$
$$\beta_{10k} = \gamma_{100} + \gamma_{101}GOV_{10k} + u_{10k}$$
$$\beta_{20k} = \gamma_{200} + \gamma_{201}IND_{20k} + u_{20k}$$
$$\beta_{30k} = \gamma_{300} + u_{30k}$$
$$\beta_{40k} = \gamma_{400} + \gamma_{401}LIKSCI\_3_{40k} + u_{40k}$$
$$\beta_{41k} = \gamma_{410} \tag{12}$$

VARIANCE EXPLAINED

The concept of variance explained is very common in multiple regression analysis. It gives the idea of how much of the variability of the dependent variable is accounted for by linear regression on the predictor variables. The usual measure of the proportion of variance explained is the square multiple correlation, R^2. One way to approach this concept is to treat separately proportional reductions in the estimated variance components, σ^2, τ_0^2, and φ_0^2 at Level 1, 2, and 3 respectively as analogues of R^2 values at each level.

Variance components for the null model: σ_n^2, τ_{n0}^2, and φ_{n0}^2.

Variance components for the final model: σ_f^2, τ_{f0}^2, and φ_{f0}^2.

Proportion of variance explained at each level in the final model:

At Level 1: $$R_1^2 = \frac{\sigma_n^2 - \sigma_f^2}{\sigma_n^2}$$

At Level 2: $$R_2^2 = \frac{\tau_{n0}^2 - \tau_{f0}^2}{\tau_{n0}^2}$$

At Level 3: $$R_3^2 = \frac{\varphi_{n0}^2 - \varphi_{f0}^2}{\varphi_{n0}^2} \qquad (13)$$

However, this approach can be somewhat problematic. It sometimes happens that adding explanatory variables increases rather than decreases some of the variance components. Therefore, it is possible to obtain negative values of R^2. Snijders and Bosker (1999) gave a suitable multilevel version of R^2 for the two-level model where the average class size was n_2 as follows:

At Level 1: $$R_1^2 = 1 - \frac{\sigma_f^2 + \tau_{f0}^2}{\sigma_n^2 + \tau_{n0}^2}$$

At Level 2: $$R_2^2 = 1 - \frac{\sigma_f^2 / n_2 + \tau_{f0}^2}{\sigma_n^2 / n_2 + \tau_{n0}^2} \qquad (14)$$

Equation (8) can be extended to a three-level model where on average each school consists of n_3 classrooms.

At Level 1: $$R_1^2 = 1 - \frac{\sigma_f^2 + \tau_{f0}^2 + \varphi_{f0}^2}{\sigma_n^2 + \tau_{n0}^2 + \varphi_{n0}^2}$$

At Level 2: $$R_2^2 = 1 - \frac{\sigma_f^2 / n_2 + \tau_{f0}^2 + \varphi_{f0}^2}{\sigma_n^2 / n_2 + \tau_{n0}^2 + \varphi_{n0}^2}$$

At Level 3: $$R_3^2 = 1 - \frac{\sigma_f^2 /(n_2 * n_3) + \tau_{f0}^2 / n_3 + \varphi_{f0}^2}{\sigma_n^2 (n_2 * n_3) * + \tau_{n0}^2 / n_3 + \varphi_{n0}^2} \qquad (15)$$

Variance components presented in Table 3 were calculated using equation (15).

RESULTS

The Null Model: Differences Between Schools and Between Classes

The analysis was started by fitting the null model. This model provides estimates of the differences between students, between classes and between schools. The sum of these three components is the total variance. It can be seen in Table 3 that for science achievement, 53.3 per cent (38.07/71.45) of the total variance is situated at the student level and another 46.6 per cent (33.34/71.45) of the total variance is located at the class level. These large components indicate that there are large differences between students and between classrooms. The percentage of the variance at the school level is very small (0.04/71.45=0.1%) which suggests that the schools are very similar to each other in terms of student achievement in science. In other words, the Level 3 intraclass correlation expressing the likeness of students in the same school is estimated to be 0.001, while the intraclass correlation expressing the likeness of students in the same classes and the same schools is estimated to be 0.47. Since most of the variance components at the school and class levels are situated at the class level, it is important to localise school performance to a specific classroom or grade level.

Type A Model: Adding Student Characteristics

At the student-level, the results in Table 2 show that Science achievement is directly influenced by Father's occupation (FOCC), Expected occupation (EXPED), Like science (LIKSCI) and Prior achievement (PRIORACH). When other factors were equal, students whose fathers had high status occupations outperformed students whose fathers had low status occupations. Students who aspired to pursue education to higher 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.

Adding the student level variables to the model explains a large part of the differences between students (52.7 %), classes (69.9 %), and between schools (69.8 %) in science achievement. In other words, science achievement differences between schools and between classes were largely due to intake differences at the grade level under survey. The remaining differences between classes and between schools were indicators of the variance in Type A school effects and in Type A teaching effects. The residuals of schools and classes can be seen in Figure 2 and Figure 3 respectively, with little variability between schools.

Type B Model: Adding Classroom and School Contexts

From Table 2 it can be seen that at the class-level, Science achievement is directly influenced by Average prior achievement (PRIOR_2) and Class size (CSIZE). When other factors were kept 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. Nevertheless, in interpreting the effects of class size, it needs to be recognised 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 is not surprising. Students in the schools that implemented streaming policy achieved better in science.

Table 2. Final estimation of fixed effects

Fixed Effects		Type A model Coefficient	S.E	p-value	Type B model Coefficient	S.E	p-value	Type X model Coefficient	S.E	p-value	Type Z model Coefficient	S.E	p-value
Intercept	γ_{000}	28.52	0.25	0.000	28.34	0.31	0.000	28.31	0.32	0.000	27.17	0.50	0.000
STREAM	γ_{001}										1.62	0.59	0.017
PRIOR_2	γ_{010}				0.27	0.08	0.003	0.28	0.07	0.002	0.29	0.06	0.001
CSIZE	γ_{020}				0.28	0.06	0.001	0.29	0.07	0.001	0.30	0.06	0.000
FOCC,	γ_{100}	0.37	0.12	0.011	0.37	0.14	0.016	0.68	0.19	0.004	0.71	0.19	0.003
Interaction with GOV	γ_{101}							-0.50	0.22	0.044	-0.55	0.22	0.029
EXPED	γ_{200}	0.58	0.09	0.000	0.50	0.10	0.000	0.44	0.10	0.000	0.43	0.10	0.001
Interaction with IND	γ_{201}							0.54	0.28	0.072	0.66	0.29	0.041
LIKSCI	γ_{300}	0.14	0.01	0.000	0.15	0.01	0.000	0.14	0.01	0.000	0.15	0.01	0.013
PRIORACH	γ_{400}	0.97	0.05	0.000	0.93	0.04	0.000	0.94	0.04	0.000	0.94	0.04	0.000
Interaction with LIKSCI_3	γ_{401}				0.01	0.00	0.024	0.01	0.00	0.027	0.01	0.00	0.028
Interaction with FOCC_2	γ_{410}				0.07	0.02	0.020	0.05	0.02	0.023	0.06	0.02	0.026

Table 3. Variance Components

Model	Deviance	Number of Parameter Estimated	**Available**			**Explained (%)**			**Unexplained (%)**		
			Student (N=1984)	**Class** (K=71)	**School** (J=15)	**Student** (N=1984)	**Class** (K=71)	**School** (J=15)	**Student** (N=1984)	**Class** (K=71)	**School** (J=15)
Null model	13,078	4	38.07	33.34	0.04						
Prior Achievement	12,142	9	24.22	9.58	0.02	52.7	69.9	69.8	47.3	30.1	30.2
Type A Model	11,879	36	20.68	6.49	0.14	61.8	78.8	77.4	38.2	21.2	22.6
Type B Model	11,792	61	20.56	1.63	0.81	67.9	90.9	82.2	32.2	9.1	17.8
Type X Model	11,786	55	20.54	1.63	0.71	67.8	91.1	83.6	32.1	8.9	16.4
Type Z Model	11,783	56	20.54	1.74	0.31	68.4	92.0	88.7	31.6	8.0	11.3

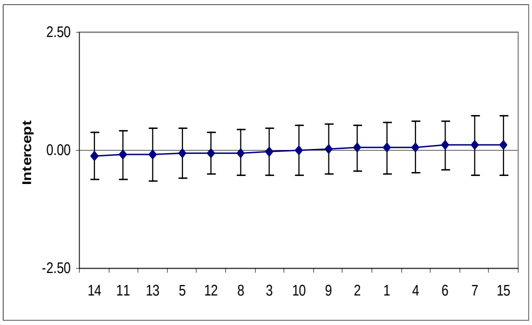

Figure 2. Type A school residuals

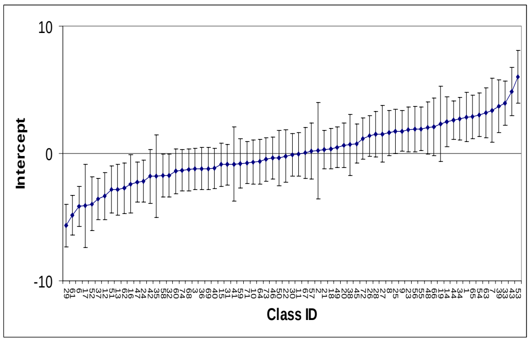

Figure 3. Type A class residuals

In Table 2 there are two significant cross-level interaction effects. These cross-level interaction effects are between (a) PRIORACH and FOOC_2 at Level 2 (class level); and (b) PRIORACH and LIKSCI_3 at Level 3 (school level). It can be seen in Figure 4 and Figure 5 that the effect of prior achievement is stronger in classes with higher status of fathers' occupation and in schools with higher level of liking science. Higher achieving students were better off in classes that had higher status of fathers' occupation as well as in schools with higher levels of liking science.

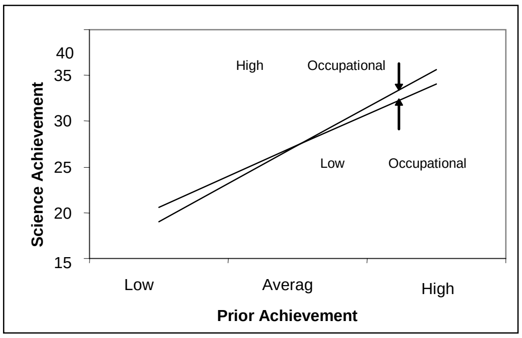

Figure 4. Impact of interaction effect of FOCC_2 and PRIORACH on Science Achievement at the classroom level

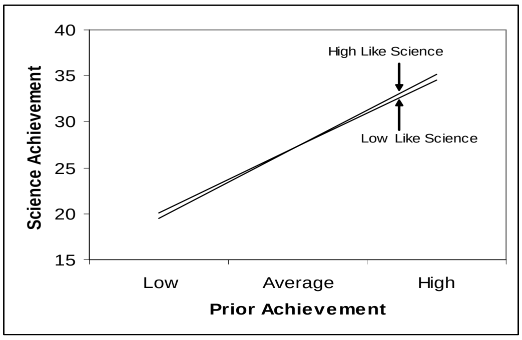

Figure 5. Impact of interaction effect of LIKSCI_3 and PRIORACH on Science Achievement at the school level

After controlling for student characteristics, class context and school context, the proportion of variance explained is increased by 9.1 per cent at the student level, 8.9 per cent at the class level, and 7.6 per cent at the school level. The residuals of 15 schools and 71 classes can be seen in Figure 6 and Figure 7, with an increase in variability between schools and a decrease in the variability between classes.

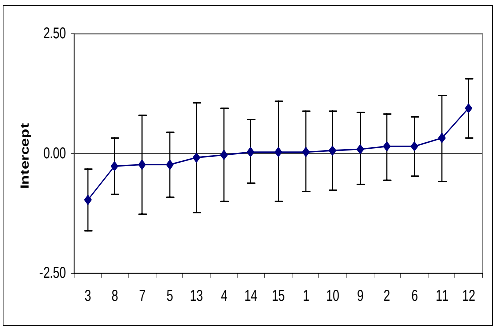

Figure 6. Type B school residuals

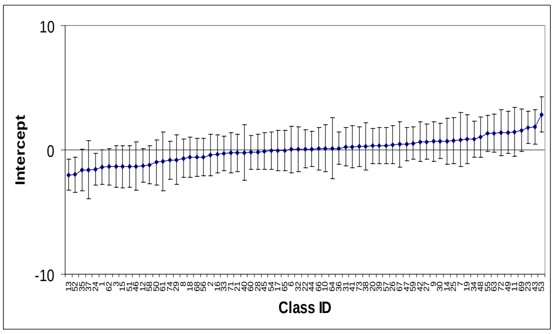

Figure 7. Type B class residuals

Type X Model: Adding Non-Malleable School Policies and Practices

When non-malleable policies were entered into the equations at Level 3, two additional interaction effects were found. These interaction effects included interactions between (a) FOCC and GOV and (b) EXPED and IND. From Figure 8 it can be seen that when other factors are equal, father's occupation had less impact in government schools than in non-government schools. In other words, students with high father's occupational status gained smaller advantage in government schools compared non-government schools.

Likewise, from Figure 9 it can be seen that students in independent schools achieve higher scores in science when they have high expected education. However, students with low levels of expected education have noticeably lower levels of achievement if they are in independent schools.

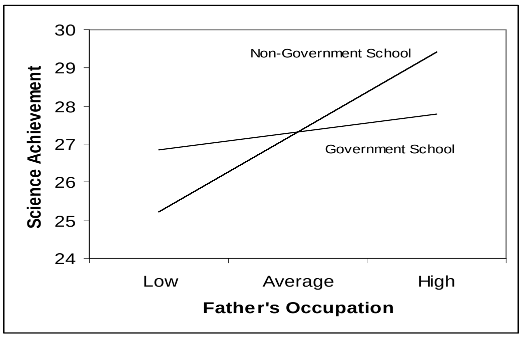

Figure 8. Impact of interaction effect of Government School and FOCC on Science Achievement at the school level

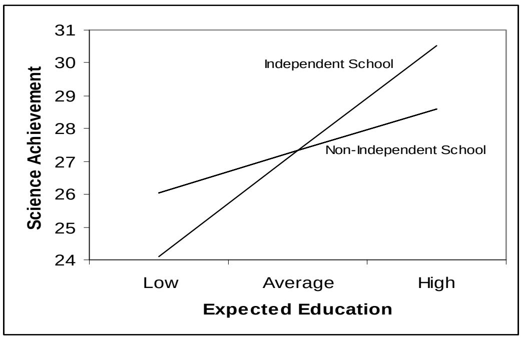

Figure 9. Impact of interaction effect of Independent School and EXPED on Science Achievement at the school level

After adding non-malleable policies and practices, only 16.4 per cent and 8.9 per cent of variance components at the school and class levels are left unexplained. The Type X residuals of 15 schools and 71 classes can be seen in Figure 10 and Figure 11 respectively.

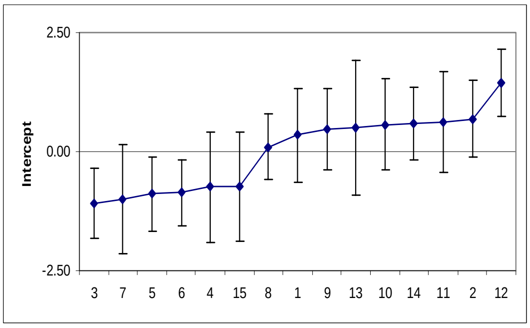

Figure 10. Type X school residuals

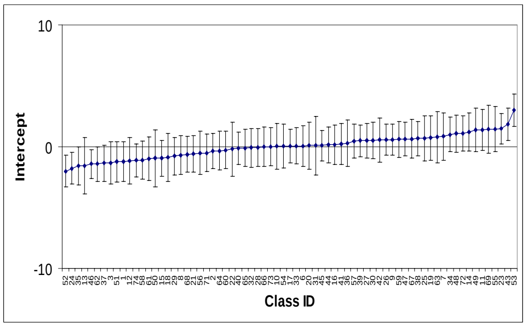

Figure 11. Type X class residuals

Type Z Model: Adding Malleable School Policies and Practices

At the School level, the results in Table 2 show that Science Achievement is also directly influenced by streaming policy (STREAM). Students in the schools that implemented streaming policy achieved better in science. In this model, only 31.6 per cent, 8.9 per cent and 11.3 per cent of variance components at student, class, and school levels are left unexplained. The Type Z residuals of 15 schools and 71 classes can be seen in Figure 12 and Figure 13 respectively.

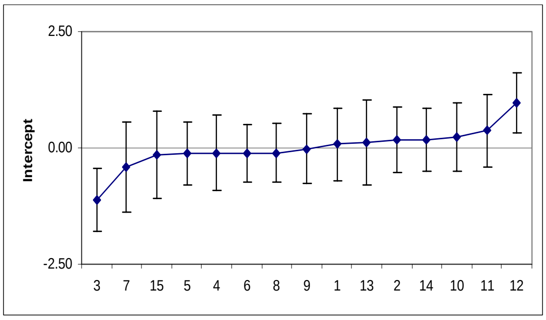

Figure 12. Type Z school residuals

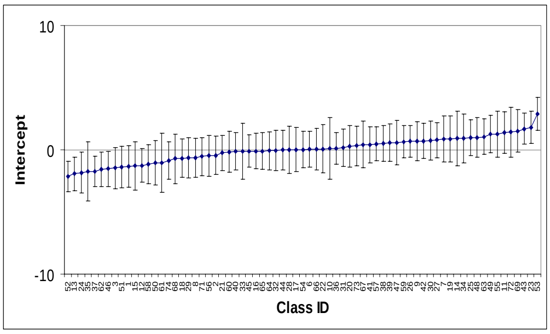

Figure 13. Type Z class residuals

Initially the differences between schools are very small as shown by the residuals of the Null model. After controlling for student characteristics, there are still no significant differences between schools. Adding classroom context and school context variables noticeably change the residuals for School 3 and School 12. Making allowance for additional non-malleable policies changed school residuals even further. However, after controlling for the significant malleable policy variable the average levels of performance for most schools are not significantly different from each other. School 3 and School 12 are the two schools that have noticeably lower and higher performance respectively. School 3 is significantly worse than other schools, but School 12 is significantly better than other schools after controlling for student characteristics, context variables as well as identified school policies and practices. These changes are noticeable from comparison of Figures 2, 6, 10, and 12 as well as from Figure 14, after allowance is made for the Type A, Type B, Type X and Type Z effects.

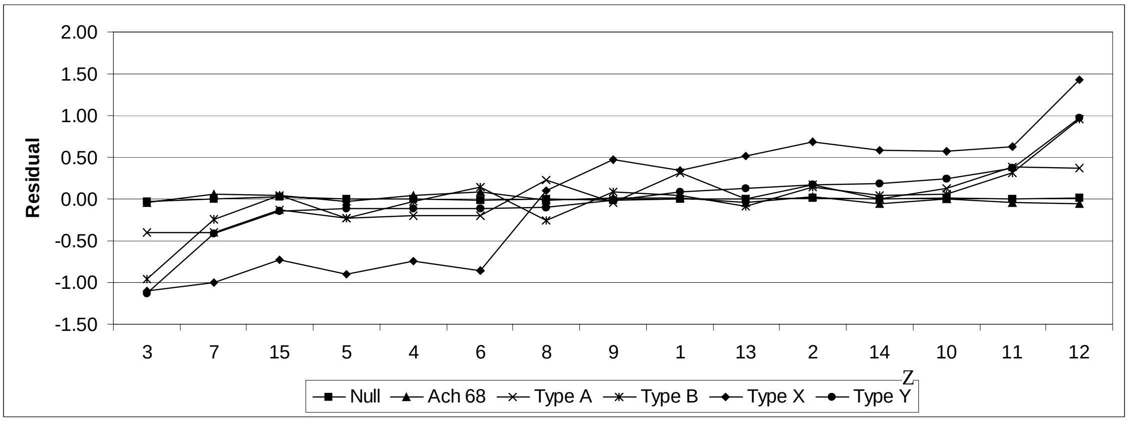

Figure 14. Types A, B, X and Z school residuals

CONCLUSIONS

This article is concerned with the statement that, student outcomes are only partially influenced by the school where they are enrolled. Other factors that have an impact on the student outcomes are student characteristics and context variables. In this study, Type A, Type B, Type X and Type Z effects are estimated by allowing for student background, class and school contexts, non-malleable school policies and malleable school policies respectively in successive regression equations.

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 32 per cent of the variance between students within classrooms is left unexplained, indicating that there are other student-level factors likely to be involved in influencing student achievement.

At the classroom level, about eight per cent of the total classroom variance or only 1.7 per cent of the total variance is left unexplained, with the average level of prior achievement of the class group having a significant effect. In addition, class size has a positive effect at this level on science achievement, with students in larger classes doing significantly better than students in smaller classes. This effect is likely to be confounded with factors associated with the qualities of the teachers assigned to teach the larger and the smaller class groups. Perhaps, this indicates the skill of the administration of the schools, particularly in those schools that adopt streaming practices to select the best teachers and allocate them to the higher performing students in larger classes. In addition, an interaction effect also reveals that the effect of prior achievement is stronger in classes with high status of fathers' occupation. High achieving students are better off in classes that have higher status of fathers' occupation.

At the school level of analysis, streaming directly explains some of the differences in levels of performance between schools in spite of the very small between school variance. The influences of the non-malleable variables involving school type and whether a school is single-sex or coeducational do not have direct effects on the educational outcome of science achievement, but they do have moderating or interaction effects. Thus whether the school is a Government or an Independent school interacts with Father's occupation and Expected education respectively to have small effects on the outcome variable. Nevertheless, it is this factor of school type that has had and continues to have a marked influence on changes in the provision of education in the Australian school systems. Unfortunately it is no longer possible to undertake research into this issue, because over-simplistic value added comparisons, that were made prior to the introduction of multilevel analytical procedures have contaminated this field of inquiry in Australia.

Two important findings emerge from this study. First, considerable variance is situated at the class level. Therefore in examining value added across schools, the class level can not be ignored. Otherwise, the class level variance components may be confounded with student level and school level variance components and lead to an overestimation of school differences. In educational effectiveness research, neglecting class context variables may lead to incorrect conclusions. Second, very little variance is left unexplained at the school and class levels to be accounted for by characteristics associated with school resources or by the direct effects of teachers. If the qualities of teachers are having effects they are associated with and are subordinate to the levels of initial achievement of the students whom they are assigned to teach, with high achieving students being placed in larger classes possibly with the better teachers.

However, the use of a value added approach in assessing school effectiveness is not without problems. There is still room for argument whether Type A, Type B, Type X or Type Z effects should be considered. Careful thought also needs to be given when considering which of the variables should be used in estimating Type A, Type B, Type X and Type Z effects. Moreover, how to obtain information on classroom effects is yet another question to address. Should longitudinal data rather than cross-sectional data be used? Apart from these problems which still need to be debated, the value added approach is providing a way to assess better the effectiveness and the accountability of schools as well as classrooms and teachers. Furthermore, it is clearly inappropriate to rank schools on terms of their performance and indeed to rank countries, without giving some consideration to these complex statistical problems. Nonetheless, research and scholarly debate needs to be carried out to develop a better understanding of the issues addressed.

REFERENCES

Ballou, D., Sanders, W. and Wright, P. (2004) Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-65.

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

De Fraine, B., Van Damme, J. and Onghena, P. (2002) Accountability of School and Teachers:what should be taken into account? *European Educational Journal*, 1(3), 403-428.

Hungi, N. (2003) *Measuring School Effects Across Grades.* Flinders University Institute of International Education Research Collection. No.6. Adelaide: Shannon Research Press.

Keeves, J.P., Hungi, N. and Afrassa, T. (2005). Measuring value added effects across schools: should schools be compared in performance? *Studies in Educational Evaluation*, 31(2-3), 247-266.

Meyer, R.H. (2000) Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies. *NISE Brief*, 3(3), 1-7.

Meyer, R.H. (2002) *Value-Added Indicators: Do They Make an Important Difference? Evidence from the Milwaukee Public Schools*, Paper presented at the Annual Meeting of the American Educational Research Association, April 2, 2002: New Orleans.

Raudenbush, S.W. and Willms J.D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307-335.

Raudenbush, S.W. and Bryk, A.S. (1994). Hierarchical linear models. In T. Husén and T.N. Postlethwaite (Eds.), *International Encyclopedia of Education: Research and Studies* (2nd ed., pp. 2590-2596). Oxford: Pergamon Press.

Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F. and Congdon, R.T. (2004). *HLM 6: Hierarchical Linear and Nonlinear Modeling* . Lincolnwood, IL: Scientific Software International.

Rowe, K.J. (2000) Assessment, league tables and school effectiveness: Consider the issues and 'let's get real'!, *Journal of Educational Enquiry*, 1(1), 73-98.

Rubin, D.B., Stuart, E.A. and Zanutto, E.L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*(1), 103-116

Snijders, T. and Bosker, R. (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. London: Sage Publications.

Willms, J.D. and Raudenbush, S.W. (1989) A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement,* 2(6), 209-232.

❖IEJ