The Influence of Secondary and Junior Schools on Sixteen Year Examination Performance: A Cross-classified Multilevel Analysis*

Harvey Goldstein and Pamela Sammons
Institute of Education, University of London

ABSTRACT

This paper presents a model for the explanation of examination (GCSE) performance at the age of 16 years in terms of both secondary and junior school attended together with prior achievement measures and certain background factors. Using a cohort of 758 students in 48 junior schools and 116 secondary schools it compares the variation in performance due to secondary schools with that due to junior schools in a multilevel cross-classified analysis. It shows that the variation among junior schools is substantially larger than that among secondary schools. It also demonstrates that those junior schools with high average achievement scores for the students when they leave junior school also tend to have high average scores for their students at the age of 16. The implications of these findings, if replicated, are profound. They imply that current attempts to measure the 'effectiveness' of secondary schools using achievement measured at the start of secondary schooling may be fruitless and they point to the need for school effectiveness research to become involved in very long term studies of schooling, rather than being restricted to a single phase.

INTRODUCTION

Increasing academic and policy interest in the issue of school effectiveness and the related concept of value-added has been evident during the last 15 years. Significant methodological advances (particularly the development of multilevel models) have improved the ways in which school

effects can be conceptualised and measured (Goldstein, 1987, 1995) and attention has been paid to the impact of issues such as differential effectiveness (Nuttall, Goldstein, Prosser, & Rashbash, 1989; Jesson & Gray, 1991; Sammons et al., 1993a; Goldstein et al., 1993), and of stability and consistency in effectiveness (Goldstein et al., 1993; Sammons et al., 1993b; Thomas et al., 1995). A growing realisation of the need to examine school class and departmental levels in greater detail (FitzGibbon, 1992; Sammons et al., 1995) is also evident. By contrast, however, scant attention has been paid to the question of the continuity of school effects measured at different stages of a student's school career. In other words, what long-term effects (if any) does previous institutional membership (e.g., primary school attended) have on later school performance at secondary school?

Sammons, Nuttall, Cuttance, and Thomas (1995) provide an example of the first attempt to address the issue of continuity in school effects using multilevel approaches. However, this study was limited in that it did not consider the full cross-classification of individual students in terms of their Secondary by Junior school attendance using recently available techniques (Goldstein, 1995). In this paper a reanalysis of this work is presented which provides a more detailed investigation of the question of continuity of school effects, estimating the joint contributions of primary and secondary schools.

Sammons et al. (1995) presents data on GCSE results on the children of the Junior school Project (JSP: Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988) followed up to age 16. They carry out two types of 2-level analysis, one with students classified by their Secondary school and one with students classified by their Junior school. They include the London Reading Test score, VR band, free school meals, social class and ethnic origin. The level 2 variance is approximately the same in both analyses.

This paper replicates that analysis (omitting some of the explanatory variables) but with both the Junior school and Secondary school identified in the same analysis as two random cross classified factors. The basic model is

$$y_{ij_1 j_2} = \sum_k \beta_k x_{kij} + u_{j_1} + u_{j_2} + e_{ij}$$

$$\mathrm{var}(u_{j_1}) = \sigma_{u_1}^2, \quad \mathrm{var}(u_{j_2}) = \sigma_{u_2}^2, \quad \mathrm{var}(e_{ij}) = \sigma_e^2, \tag{1}$$

so that the total level 2 variance is the sum of a between-Junior and a between-Secondary variance. We can elaborate this model by adding random coefficients varying at the secondary or Junior level or both. The sample size in this analysis is only 758 students with 48 Junior and 116

Secondary schools so that we should treat carefully results where there is a lack of statistical significance or lack of variation. The subscript 1 refers to Junior and 2 refers to Secondary.

The following table gives results for fitting this model with different fixed coefficients and assumptions about the level 1 variance.

We see that the between-Junior variance is always larger than that between Secondary schools. As further explanatory variables are incorporated which measure achievement at the end of Junior schooling, so relatively more of the Junior school variance is explained, as expected. From one point of view, to judge the relative contributions of Junior and Secondary school, we should not adjust for such variables as in C, in

Table 1. Variance Components Cross-classified Model for Exam Score as Response.

	A		B		C	
Fixed						
Intercept	0.51		0.50		0.25	
Males	−0.21	(0.06)	−0.19	(0.06)	−0.34	(0.07)
Free school Meal	−0.22	(0.06)	−0.23	(0.06)	−0.37	(0.08)
VR2 band	−0.39	(0.08)	−0.38	(0.08)		
VR3 band	−0.71	(0.13)	−0.71	(0.13)		
LRT score	0.31	(0.04)	0.32	(0.04)		
Random						
Level 2:						
(Junior) σ^2_{u1}	0.025 (0.013)		0.036 (0.017)		0.054 (0.024)	
(Secondary) σ^2_{u2}	0.016 (0.014)		0.014 (0.014)		0.019 (0.02)	
Level 1:						
σ^2_{e0}	0.50	(0.06)	0.554	(0.06)	0.74	(0.05)
σ_{e01}	0.092 (0.03)		0.064	(0.03)	0.10	(0.05)
σ_{e02}	0.093 (0.018)					
σ^2_{e2}	0.033 (0.022)					
−2Log likelihood	1848.8		1884.2		2130.3	

Note. The exam score and LRT score have been transformed empirically to have N(0,1) distributions. FSM is a binary (yes, no) variable. At level 2 the subscript 1 refers to Junior and 2 to secondary school. At level 1 the subscript 0 refers to the intercept, 1 to males and 2 to LRT.

which case the Junior schools exhibit three times the variability of secondary schools. Partly, presumably this is because they are smaller, but may also reflect the importance of early schooling.

From another point of view, analyses A or B (the former is a better specified model) of Table 1 are the most useful, since they present a measure of the 'value added' by secondary schools after adjusting for intake performance *and* the effect of the Junior school attended which is not captured by the LRT and VR band variables. Thus, although analysis A or B is the most relevant from this viewpoint, we cannot directly compare the Junior and Secondary variances, because end-of Junior attainments have been fitted. We note that these results are broadly in line with those reported for Scottish data from schools in Fife (Goldstein, 1995). Finally we look at models that additionally fit LRT at level 2.

Table 2. Table 1 with LRT Random at Higher Levels.

	A		B	
Fixed				
Intercept	0.49		0.51	
Males	−0.21	(0.06)	−0.21	(0.06)
Free school Meal	−0.22	(0.06)	−0.22	(0.06)
VR2 band	−0.36	(0.08)	−0.39	(0.09)
VR3 band	−0.69	(0.13)	−0.71	(0.13)
LRT score	0.33	(0.04)	0.31	(0.04)
Random				
Level 2:				
(Junior) $\sigma^2_{u1(0)}$	0.027 (0.014)		0.025 (0.013)	
(Secondary) σ^2_{u2}	0.028 (0.017)		0.016 (0.014)	
$\sigma_{u2(01)}$	0.028 (0.010)		0	
$\sigma^2_{u2(1)}$	0.009 (0.011)		0	
Level 1:				
σ^2_{e0}	0.55	(0.041)	0.50	(0.04)
σ_{e01}	0.056 (0.032)		0.091 (0.059)	
σ_{e02}			0.094 (0.020)	
σ^2_{e2}			0.033 (0.022)	
-2Log likelihood	1874.9		1848.8	

It was not possible to fit LRT random across Junior schools. The interpretation of such a model would have been that the contribution of Junior schools was 'differential' that is that it was a function of the LRT score. This would be difficult to interpret.

There is significant random coefficient variation, between secondary schools, for LRT. This is in contrast to the results in Sammons et al. (1994) where there was a significant coefficient for LRT for Junior but not Secondary schools, although this may in part be due to the fact that the latter authors were working with the original score scales for the exam score and LRT. Note in analysis A, however, that the large covariance implies a correlation greater than 1.0, but, more importantly, predicts a negative between-secondary school variance for values of (Normal score) LRT below about −0.45. This implies an imperfectly specified model. The existence of such a negative variance is acceptable, so long as the total level 2 variance is positive, but for small enough values this too will be violated. One solution is to make the between-Secondary variance a nonlinear (exponential) function of LRT and this can be done using the software MLn. Essentially there will be (low) values of LRT for which there is only variation due to Junior schools. Analysis B, however, which is the same as A in Table 1, shows that when the level 1 variation is fully specified the level 2 random coefficient for LRT disappears. This underlies the importance of properly specifying the level 1 variation.

When we fit LRT random for Secondary schools at level 2 we should also fit LRT in the fixed part of the model. If we do not do this we will obtain a spuriously high between-Secondary variance for this coefficient because the model will estimate this variation about zero rather than the average LRT coefficient of 0.31. Thus, we cannot directly compare the Junior and Secondary variances, as discussed above, when LRT is random at level 2.

In the final analyses we fit the Maths and English test scores obtained on entry to Junior school. There is little overall effect of either Maths or English. Adjusting for these does not change the level 2 variances when the 11-year LRT and VR band variables are included. As expected, however, when the 11-year variables are excluded in analysis C, the between-Junior variance is reduced to about three quarters the value in analysis C of Table 1, but the between-Secondary variance is altered little. Thus, once we have taken account of intake differences at Junior school, the between-Junior variance is more than twice as great as that between Secondary schools. Analysis B adds the LRT score at level 1 and this reduces the between-Junior school variance by a further quarter without changing that between Secondary schools appreciably.

Table 3. As Table 1 with 8-year-old Maths and English test scores measured about values close to their means; 25 and 50 respectively.

	A		B		C	
Fixed						
Intercept	0.15		−0.99	(0.12)	−1.08	
Males	−0.22	(0.06)	−0.31	(0.06)	−0.27	(0.07)
Free school Meal	−0.22	(0.06)	−0.25	(0.07)	−0.25	(0.07)
VR2 band	−0.36	(0.09)				
VR3 band	−0.66	(0.14)				
LRT score	0.29	(0.05)				
8-year English score	0.00016	(0.0020)	0.011	(0.002)	0.011	(0.002)
8-year Maths score	0.0058	(0.0056)	0.026	(0.006)	0.028	(0.006)
Random						
Level 2:						
(Junior) σ^2_{u1}	0.025	(0.014)	0.030	(0.016)	0.040	(0.018)
(Secondary) σ^2_{u2}	0.016	(0.014)	0.020	(0.016)	0.017	(0.016)
Level 1:						
σ^2_{e0}	0.50	(0.040)	0.54	(0.049)	0.60	(0.04)
σ_{e01}	0.093	(0.030)	0.10	(0.04)	0.08	(0.04)
σ_{e02}	0.093	(0.018)	0.11	(0.04)		
σ^2_{e2}	0.031	(0.021)	0.07	(0.03)		
-2 Log likelihood	1847.6		1949.9		1964.8	

It seems, on the basis of these analyses that the best value added measure for secondary schools is that of analysis A of Table 3 where the between-Secondary variance is only 0.6. For comparison, if we retain the fixed part of this model but adjust for 11-year variables only and exclude the Junior school classification, we obtain the results in Table 4. Adding the eight year Maths and English test scores has a negligible effect.

Analysis A in Table 4 is the 'standard' Secondary school effectiveness model, and we can see that it 'overestimates' the Secondary school effect, in this case producing a variance that is considerably larger than the estimate when the Junior school variation is included. Comparing the

Table 4. As Analysis B of Table 2. A: Junior School Classification not Fitted; B: Crossing Fitted as a Single Level 2 Classification with 286 Units.

	A		B	
Fixed				
Intercept	0.50		0.51	
Males	−0.22	(0.06)	−0.22	(0.06)
Free school Meal	−0.22	(0.06)	−0.21	(0.06)
VR2 band	−0.37	(0.09)	−0.38	(0.09)
VR3 band	−0.65	(0.13)	−0.67	(0.13)
LRT score	0.32	(0.04)	0.32	(0.04)
Random				
Level 2:				
(Secondary) σ_{u2}^2	0.028 (0.015)		0.032 (0.019)	
Level 1:				
σ_{e0}^2	0.52	(0.04)	0.52	(0.04)
σ_{e01}	0.08	(0.03)	0.08	(0.03)
σ_{e02}	0.10	(0.02)	0.10	(0.02)
σ_{e2}^2	0.03	(0.02)	0.03	(0.02)
-2 Log likelihood	1852.8		1855.1	

deviances we also that there is a significant difference between the fits at 5 per cent. For the second analysis, where the complete cross classification is treated as a single level 2 unit we obtain very similar results with a non significant change in the deviance statistic. We see, therefore, that by structuring the cross classification so that the total variance is an additive function of the Junior and Secondary school variances, we obtain a more powerful test for the Junior school effect.

We have not calculated residuals, which would normally be used as estimates of school effectiveness. The relatively small numbers of students in Secondary schools implies that the residuals have very large standard errors attached to them. In certain cases we would expect the estimates from the two analyses to be similar, in particular where for each secondary school the distribution of students with respect to the primary schools are similar. In general, however, the residuals from these two analyses would not be expected to rank schools in the same order.

We conclude therefore that the usual quantitative procedures for estimating school effectiveness need to be augmented with careful measurements of all relevant prior performances, including institutional membership. This also applies to studies of value added at A level, where, in principle, we can study the variation from Primary, secondary and tertiary institution simultaneously. This analysis has demonstrated that quantitative studies of school effectiveness need to turn their attention to collecting longitudinal data which will make this possible.

A BIVARIATE MODEL

We now extend the above models to consider the joint modelling of the GCSE score and the reading test score at the end of Junior school. In principle we could extend this analysis by considering the verbal reasoning band assignment as a further Junior school outcome, but we shall not consider this case. By modelling these outcomes jointly we can incorporate two Junior school 'effects', that which is associated with achievement at the end of Junior school and that which is associated with GCSE achievement. It is particularly interesting to study the relationship between these two effects and the extent to which they may be explained by further factors. This further leads us to extend the usual definition of 'school effectiveness' to include the effect of an institution on progress in subsequent institutions.

The basic model can be written as follows, extending the notation in equation (1)

$$y_{ij_1 j_2}^{(2)} = \sum_k \beta_k^{(2)} x_{kij}^{(2)} + u_{j_1}^{(2)} + u_{j_2} + e_{ij}^{(2)}$$

$$y_{ij_1}^{(1)} = \sum_k \beta_k^{(1)} x_{kij}^{(1)} + u_{j_1}^{(1)} + e_{ij}^{(1)} \tag{2}$$

$$\text{cov}(e_{ij}^{(1)} e_{ij}^{(2)}) = \sigma_{e_1 e_2}, \quad \text{cov}(u_{j_1}^{(1)} u_{j_1}^{(2)}) = \sigma_{u_1 u_2}.$$

In our models the Junior school response variable $y_{ij_1}^{(1)}$ will also appear as an explanatory variable for the Secondary school response $y_{ij_1 j_2}^{(2)}$. In this case (2) is to be interpreted as a conditional rather than unconditional path model. We note, however that in this case, where it is secondary school progress which is of interest, by conditioning on 11 year achievement we would expect this to remove the correlation between the Junior school effects in the two components of the model. For the same reason,

Table 5. Bivariate response model for GCSE and LRT scores.

	A	B	C
Fixed:	Estimate (s.e.)	Estimate (s.e.)	Estimate (s.e.)
GCSE response:			
Intercept	0.48	0.62	0.08
Males	−0.24 (0.07)	−0.29 (0.05)	−0.31 (0.07)
Free school meals	−0.21 (0.08)	−0.20 (0.08)	−0.21 (0.08)
LRT score	0.33 (0.06)	—	—
VR2 band	−0.33 (0.10)	−0.45 (0.10)	—
VR3 band	−0.67 (0.17)	−0.95 (0.16)	—
8-year English	−0.001 (0.002)	0.006 (0.002)	0.012 (0.002)
8-year Maths	0.006 (0.007)	0.013 (0.007)	0.027 (0.007)
LRT response:			
Intercept	0.08	0.08	0.08
Gender	−0.15 (0.05)	−0.15 (0.05)	−0.14 (0.05)
Free school meals	−0.05 (0.06)	−0.05 (0.06)	−0.05 (0.06)
8-year English	0.023 (0.001)	0.023 (0.001)	0.023 (0.001)
8-year Maths	0.037 (0.005)	0.037 (0.005)	0.036 (0.005)
Random:			
Between School:			
$\sigma^2_{u_2}$	0.009 (0.018)	0.009 (0.018)	0.006 (0.017)
$\sigma^2_{u_1}(2)$	0.036 (0.020)	0.043 (0.022)	0.037 (0.021)
$\sigma^2_{u_1}(1)$	0.055 (0.018)	0.055 (0.018)	0.055 (0.018)
$\sigma_{u_1 u_2}$	0.004 (0.013)	0.024 (0.014)	0.036 (0.014)
Between Student:			
$\sigma^2_{e_0}(2)$	0.634 (0.038)	0.660 (0.039)	0.712 (0.042)
$\sigma_{e_0}(1)$	0.366 (0.021)	0.366 (0.021)	0.366 (0.021)
$\sigma_{e_0}(1,2)$	−0.058 (0.040)	0.048 (0.041)	0.084 (0.041)
-2 Log likelihood	2767.7	2792.2	2825.9

Note. Superscript 1 refers to LRT as response, superscript 2 to GCSE as response: sub-
script 1 refers to males, 2 to LRT and 0 to the intercept for the complex between-
student variation.

if the model is well specified it will also remove the correlation between the student level residuals.

Model (2) is fitted by defining a 3-level model where 'type of response', Junior or Secondary, is level 1, student is level 2 and level 3 is the cross classification. From a computational viewpoint it is efficient to define a 4-level model where level 3 is the primary school classification, level 4 consists of a single unit with the secondary school variation defined by a set of dummy variables with coefficients random at level 4 and having a common (between-secondary) variance (Rasbash, Woodhouse, Yang, & Goldstein, 1995, Goldstein, 1995). Table 5 presents the results from fitting models based on (2). Details of the model fitting procedure are given in the appendix.

Various complex student-level models were fitted and we show just the models fitting variance components. From analysis A we see that the Secondary school variance is still small, with a large standard error, and the between Junior variance for GCSE as response has increased somewhat, as has its standard error. As expected, the correlation between the Junior school contribution to GCSE and the Junior school contribution to LRT (0.08) is very small, as is the corresponding student level correlation (–0.15).

Analysis B in Table 5 omits the 11 year LRT score with the result that the student level correlation rises to 0.10 and the Junior school correlation to 0.49. When the verbal reasoning band dummy variables are removed from the model in addition these correlations rise to 0.16 and 0.81 respectively as shown in analysis C. The simple correlation between the Normalised GCSE score and the Normalised LRT score is 0.53.

DISCUSSION

It appears that there is a strong continuity of Junior school effects for 11 year scores and GCSE scores as outcomes, where the GCSE score is not adjusted for achievement at the end of Junior schooling. It is also clear from the earlier analyses that it is the inclusion of the Junior school identification which causes the largest reduction in the Secondary school variance. In addition, the relatively small correlation between the level 1 residuals in analysis C of Table 5 suggests that the Junior school may be exerting a strong, persisting, influence on 16 year achievement. Because of the relatively small sample size in the present analyses, our conclusions must remain tentative, but they raise interesting and potentially important questions about the use of data relating only to the Secondary

school period in order to study variation between Secondary schools. In particular the results suggest that 'value added' approaches to comparing Secondary schools adjusting for intake achievement but not for previous schooling, may be seriously deficient.

Whilst our analyses have focused on the continuing impact of junior schools at GCSE, the implications of the results on the deficiencies of current models are likely to be equally applicable to analyses of perform-ance at the post 16 level. For example, proper control for intake in studies of A-level examination results for 17–18 year olds may need to take ac-count of previous institutional membership at both secondary and primary school level. The recent publication of the U.K. Schools Curriculum and Assessment Authority (SCAA) report (SCAA, 1994) on value added meas-ures of school effectiveness suffers from a number of serious limitations (see the critique by Gray, 1994). The results of our research point to addi-tional problems in this approach in view of the complexity inherent in developing models which can make appropriate adjustments for institution attended and achievement at prior stages of schooling. The need for more research into this topic is very clear and we hope to undertake further work using larger samples to explore the continuity of school effects in more depth and in a variety of socio-economic and geographical contexts.

APPENDIX

To specify a cross classified bivariate model we use level 1 to define the bivariate structure, that is with up to 2 units (the GCSE or LRT response) within each level 2 unit (student) within a cross classification of junior by secondary schools. The junior school classification is specified at level 3, and the secondary school classification at level 4, where every secondary school is assigned a dummy variable whose coefficient is random (with a single variance term) at level 4 and these variances are constrained to be equal. Because the response is bivariate at levels 2 and 3 we have the variances of GCSE and LRT and their covariance as parameters to be estimated.

This structure can be set up and estimated in Mln (Rasbash et al., 1995). Because we have 116 secondary schools, however, this demands extremely large storage requirements and the analysis can be made more tractable by omitting all cells of the cross classification with only 1 stu-dent. This yields eight disjoint 'blocks' of Secondary by Junior schools, with no more than 19 secondary schools in each block. In total this omits 31 Secondary schools, 1 Junior school and 146 students from the data set.

We have run analyses for both the full and reduced data sets, using the purely hierarchical models and there are no substantial differences in the results. The reduction in the between Secondary school variance shown in Table 5, however, is partly a result of these omissions, but there is anyway a large standard error associated with this parameter. This leads us to feel confident that we have not introduced noticeable biases by this procedure, although there is a loss of efficiency of the order of 15% in the estimation of variances at the school level.

REFERENCES

FitzGibbon, C. (1992). School effects at A level: Genesis of an information system. In D. Reynolds & P. Cuttance (Eds.), *School effectiveness research policy and practice.* London: Cassell.

Goldstein, H. (1995). *Multilevel statistical models* (2nd Ed.). London: Edward Arnold.
Goldstein, H. (1987). *Multilevel models in educational and social research.* London: Charles Griffin & Co.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Nuttall, D., & Thomas, S. (1993). A multilevel analysis of school examination results. *Oxford Review of Education, 19,* 425–433.

Gray, J. (1994). The SCAA value added report. *Times Educational Supplement,* December 16, 1994.

Jesson, D., & Gray, J. (1991). Slants on slopes: using multilevel models to investigate differential school effectiveness and its impact on pupils' examination results *School Effectiveness and School Improvement, 2,* 230–271.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters.* Wells: Open Books.

Nuttall, D, Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness, *International Journal of Educational Research, 13,* 769–776.

Rasbash, J., Woodhouse, G., Yang, M., & Goldstein, H. (1995). *Mln: software for multilevel analysis, Command reference.* London: Institute of Education.

Sammons, P., Nuttall, D., & Cuttance, P. (1993a). Differential school effectiveness: Results from a reanalysis of the Inner London Education Authority's Junior School Project data. *British Educational Research Journal, 19,* 381-405.

Sammons, P., Thomas, S., Mortimore, P., Cairns, R., & Bausor, J. (1995). *Understanding school and departmental differences in academic effectiveness.* Paper presented at the International Congress of School Effectiveness and Improvement, Leewarden Netherlands, January 1995.

Sammons, P., Mortimore, P., & Thomas, S. (1993b). Do schools perform consistently across outcomes and areas? in J. Gray, D. Reynolds, C. Fitz-Gibbon, & D. Jesson (Eds.), *Merging traditions. The future of research on school effectiveness* (Chapter 1). London: Cassell. Also paper presented to the ESRC Seminar on School Effectiveness and School Improvement, University of Steffield, July 1993.

Sammons, P., Nuttall, D., Cuttance, P., & Thomas, S. (1995). Continuity of school effects. London, Institute of Education (submitted for publication).

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Stability and consistency in secondary schools' effects on students' GCSE outcomes over three years. *School Effectiveness and School Improvement,. 8,* 169-197 (this issues).