1996 DfEE study of Value Added for 16-18 year olds in England

by

Cathal O'Donoghue (DfEE), Sally Thomas (IOE), Harvey Goldstein (IOE), and Trevor Knight (DfEE)[1]

1. Introduction

1.1 This paper examines a number of issues relevant to the modelling of value-added for GCE A/AS level students in English schools and colleges. We discuss some of the technical points involved, investigate the possibilities of estimating value-added measures for these institutions, and discuss a number of potential applications for these results by schools and colleges in advancing their effectiveness and improvement work. In this analysis, we concern ourselves only with value-added (or, progress) made by 18 year olds between GCSE and GCE A/AS Level, taking total GCE A/AS level points as our outcome variable and total and average GCSE points as our measures of prior attainment. We offer only a preliminary analysis of the large dataset at this stage. Later analysis could look at refinements to the models to consider the complex variation at the student level, and to allow for differential effects by subject of study.

1.2 Statistical modelling of value-added in education has had a high profile in recent years and has resulted in a considerable body of research. The Department for Education and Employment (DfEE) has itself carried out and commissioned some of it, including the 16-18 GCSE/GCE study by Gray and DfEE researchers (1995) using the national GCE A-level data set for 1993/94, and the study by Jesson (1996) on value-added between KS3 and GCSE in City Technology Colleges.

1.3 The Schools Curriculum and Assessment Authority (SCAA) has established a working party to advise the department on the development of a national system of value-added reporting for schools based on prior attainment, and has commissioned researchers at Durham University to carry out this work. (FitzGibbon, 1995, 1996; SCAA, 1995).

1.4 The International School Effectiveness and School Improvement Centre (ISESIC) at London University, Institute of Education has been very active in value-added research at both pre-and post-compulsory schooling, using data for Lancashire, Inner London and a national sample. (Goldstein et al., 1993; Goldstein and Thomas, 1996; Thomas and Mortimore, 1996; Thomas et al., 1995a and 1995b, Goldstein and Spiegelhalter, 1996). The National Foundation for Education Research (NFER) also has a wide range of experience of applying quantitative methods (including those used here) to the analysis of value-added for LEAs (Hutchison, 1993).

1.5 In this analysis, we continue the previous departmental research on the relationships between performance at GCSE and GCE A/AS Level. We examine how student age, gender, and institution type(s) attended impact on their progress, and consider the implications for describing this progress at the institutional level. In particular, we look at the statistical aspects of describing differences in the value-added by schools and colleges using aggregate measures of their students' performances, and then consider evidence for institutional differential effectiveness (that is, whether institutions are more, or less, effective for students of differing GCSE attainments), and how stable these institutional effects are over time.

---

[1] The authors would like to thank Denis Allnutt, John Gardner and Audrey Brown for the their helpful advice and comments throughout this project and to Mike Hart and Dave Spencer for their immense help on the IT side of things.
IOE: Institute of Education, 20 Bedford Way, London, WC1H 0AL.
DfEE: Department for Education and Employment, Sanctuary Buildings, London, SW1P 3BT

NOTE: This document contains very minor amendments in formatting and wording to the version published in the DfEE research series (March 1997).

1.6 The dataset used for this value-added analysis (described below) is hierarchical in nature, with students within cohort within institution. This type of data is most effectively analysed using Multilevel Models (MLM), a technique pioneered in school effectiveness work by Aitkin and Longford (1986) and discussed in more detail in Goldstein (1987 and 1995). Most of the analysis conducted here uses this method. However, for completeness, a number of MLM analyses have been replicated using Ordinary Least Squares regression (OLS) - of which MLM is a sophistication - in order to demonstrate any statistical differences in the results and consequent problems with interpretation which arise from using a simplified technique.

2. The Dataset

2.1 The data used in this analysis comprises the GCE A/AS level and linked GCSE results of students aged 18 in each of the three academic years ending in Summer 1993, 1994, 1995. Only those attempting at least one A /AS level examination and who had attempted one or more GCSE exams were included. A number of candidates were excluded from the analysis (Table 2.1): those who had attempted exams at one level but not both, and 'private' candidates. There were 13,160 candidates in these categories. Table 2.2 shows that over the three years there were 504,680 relevant candidates.

Table 2.1

Cases eliminated from the analysis

Category	No. of cases eliminated
Zero GCE A-level attempts	1323
Zero GCSE attempts	10761
Private Candidates	1372
Total	13160

Table 2.2

Numbers of Students and Institutions 1993-1995

	1993	1994	1995	Total
Number of students	167237	168869	168574	504680
Number of Institutions	2679	2715	2690	2824

2.2 Table 2.3 shows the scoring system assigned to each exam level. The outcome measure 'Total GCE A/AS level points' was calculated by applying the score to grades obtained and by summing over every subject attempted. 'Total GCSE points' was calculated in the same manner, and 'Average GCSE points' by dividing total points score by the number of subjects attempted at GCSE. The GCSE scores are those resulting from exams taken by the last year of compulsory schooling, typically two years before the GCE A/AS level exams. If a subject had been taken twice the highest grade was used. In addition, at A Level, 'General Studies' was omitted from our calculations because of the variation between institutions in the level of preparation they provided for their students. Finally, for ease of analysis and to reduce the effects of non-linearities, each exam aggregate was transformed to have zero mean and unit standard deviation, and 'total A/AS points' were transformed to normality. The transformations were carried out at the beginning of the analysis and variables were not restandardised for models using only subsets of the data.

Table 2.3

Scoring system used for aggregate GCSE and GCE scores.

Grade	GCSE	AS Level	A Level
A	7	5	10
B	6	4	8
C	5	3	6
D	4	2	4
E	3	1	2
F	2	0	0
G	1	0	0
U	0	0	0

2.3 Table 2.4 shows the mean and standard deviation of the attainment variables before transformation. On the basis of the scoring system given in Table 2.3, the average 'total GCE A/AS level points score' was 14, equivalent to a grade combination of CCE. The mean 'average GCSE points score' was 5.6, which indicates an average grade per subject attempted of between grades B and C. The mean 'total GCSE points score' was 50, equivalent to about 10 passes at grade C.

2.4 As all of these variables have been standardised for subsequent analyses, it is important to note their standard deviations before transformation for interpretation of the results. The standard deviation for 'total GCE A/AS Level points' was 9.73, and for 'average GCSE points' and 'total GCSE points', about 0.9 and 10 respectively. For these variables, each coefficient in the model output refers to the number of standard deviations change in 'total A/AS Level points' which would result from a change of one standard deviation in the independent variables 'total GCSE points' and 'average GCSE points'.

Table 2.4

Mean and standard deviations for GCE and GCSE scores averaged over three years

Variable	Mean	SD
Total GCE A-level	14.01	9.73
Average GCSE	5.61	0.86
Total GCSE	50.5	9.97

2.5 We now discuss the variation in GCE A-level and GCSE scores across institution type as shown in Table 2.5. This shows the mean institution-type exam score in 1994 and 1995. On average, students taking A/AS Levels in grammar schools have the highest GCSE scores, whereas the lowest mean GCSE attainments are at 'other' maintained schools - mostly secondary modern schools - and at FE colleges other than sixth form colleges. Selective independent schools, on average, have students who achieve the highest GCE A/AS Level points scores, and FE colleges the lowest, though the average in colleges will be low for many students who have taken a combination of vocational and academic subjects.

Table 2.5. Mean average exam score by institution type

Institution Type	Average GCSE points score		Total GCSE points score		Total A/AS Level points score	
	1994	1995	1994	1995	1994	1995
LEA Comprehensive	5.4	5.5	49	49	11	12
LEA Grammar	6.0	6.0	54	55	17	18
LEA Other Maintained	4.9	5.0	42	43	8	9
GM Schools	5.5	5.5	50	50	13	13
Selective independent Schools	5.9	5.9	52	52	17	18
Sixth Form College	5.5	5.6	50	50	13	14
Other FE colleges	5.0	5.0	44	44	8	8

2.6 Table 2.5 shows the broad relationship between GCSE attainment and subsequent A/AS Level achievements. At aggregate level, the higher the GCSE qualifications the higher the A/AS Level results. Table 2.6 indicates the statistical correlations between the A/AS Level and GCSE measures taken over the students in the dataset.

Table 2.6. Correlations between main variables

	Average GCSE	Total GCSE
Total GCSE	0.84	
Total GCE A-level	0.70	0.62

Average GCSE points score is slightly more highly correlated with total GCE A/AS Level scores than is total GCSE points score. The correlations are 0.70 and 0.62 respectively. Thus, overall, approximately half the variation between students in their A/AS Level points score is explained by their average GCSE points score. There will be a wide variety of personal and institutional reasons why the relationships between GCSE and A/AS Level achievements are imperfect. In what follows, we attempt to model how the power of GCSE results to statistically explain A/AS Level performance varies according to a range of institution and student-based variables, and to gauge their statistical significance.

3. The relationship between GCE A level and GCSE scores

3.1 In this section we investigate the average relationship between GCSE and GCE A/AS Level scores. Many studies, including Goldstein and Thomas (1996) and Gray et al. (1995), have shown that relationships between the measures are different at the attainment extremes. This also suggests that a linear line-of-best-fit may not best describe the overall relationship. A linear relationship may therefore not be a fair predictor of value-added for institutions with differing numbers of low and high achieving students, and we need to test whether value-added effects are smaller or greater according to the achievement-mix of students within institutions. Thus, for example, a linear model may predict negative A-level scores for students with low GCSE attainments, and the progress to A/AS Level made by very able individuals may appear to be understated.

3.2 To corroborate the technical importance of developing an appropriate non-linear relationship between the GCSE and A/AS Level attainment measures, we re-ran, using the full 1993 cohort data from our dataset, the Goldstein and Thomas (1996) (G&T) value-added models which were based on students GCSE and A/AS Level attainments from a sample of 436 institutions in England, some 15% of institutions in our dataset. As in the G&T study (and in Gray et all (1995)) we found that the statistical relationship between total GCE A-level and GCSE was improved by a high order polynomial function.

3.3. Table 3.1 shows the similarity between the coefficients estimated using our full dataset, and those computed from the G&T sample dataset. This consistency lends weight to the view that a non-linear association between GCSE and A/AS Level measure is a persistent and real feature of the data. Accordingly, the models developed in this paper fit (where statistically justified) polynomial elements in both average and total GCSE points as well as fitting a spline function for higher attaining students.

Table 3.1

Value- added coefficients from Goldstein and Thomas (1995) compared with full DfEE dataset

	DfEE model 1	G&T model 1	DfEE model 2	G&T model 2
Fixed:				
Intercept	-0.003	-0.009	0.002	-0.1
GCSE Total	0.78	0.8	0.81	0.84
GCSE 2	0.07	0.11	0.075	0.12
GCSE3	-0.042	-0.035	-0.04	-0.035
GCSE4	-0.0056	-0.005	-0.0058	-0.005
Girls	-0.059	-0.08	-0.07	-0.071
GCSE*Girls			-0.057	-0.074
GCSE2*Girls			-0.005	-0.013

3.4. We now discuss the polynomial relationship for our three year (1993-1995) dataset. Separate analyses using total and average GCSE were carried out and we report here the combined analysis where the A/AS Level score is a function of both total and average GCSE score. We used a spline function to graduate the average relationship to avoid the need to fit high order two-dimensional polynomials. For analyses which involve separate cohorts we have fitted a multi-level (or variance component) model where institution is level 3, year or cohort level 2, and student level 1 defined by:-

$$\begin{aligned} y_{ijk} = {} & \alpha + \alpha_1 x_{1ijk} + \alpha_2 x_{1ijk}^2 + \alpha_3 x_{1ijk}^3 + \alpha_4 (x_{1ijk} - t_{11})_+^3 + \alpha_5 (x_{1ijk} - t_{12})_+^3 + \\ & \beta_1 x_{2ijk} + \beta_2 x_{2ijk}^2 + \beta_3 x_{2ijk}^3 + \beta_4 (x_{2ijk} - t_{21})_+^3 + \beta_5 (x_{2ijk} - t_{22})_+^3 + \\ & v_k + u_{jk} + \varepsilon_{ijk} \end{aligned} \qquad (1)$$

where the spline or + function at the join point $z = t$ is defined as:-

$$(z - t)_+^3 = \begin{cases} (z - t)^3 & \text{if } z > t \\ 0 & \text{otherwise} \end{cases}$$

and,

x1 = Total GCSE points
x2 = Average GCSE points
v = Institution level residual with suffix k (level 3);
u = Cohort level residual with suffix jk (level 2);
e = Student level residual with suffix ijk (level 1)

producing a smooth join for the additional cubic component at $z = t$. Several other values for the join points were tried and for total GCSE score it was found that the spline components added very little power to prediction and neither did the cubic or quadratic terms and these have been omitted from the analyses reported here. For the average GCSE score it was found that a single join point at the value of 1.0 appeared to be adequate[2].

[2] Note that in the present analysis the total and average GCSE point scores are *standardised* whereas the total A/AS level points score is standardised and normalised.

3.5 Table 3.2 shows the results of fitting three models:-

1. an 'intercept' model (3) which provides an initial description of the partition of total A/AS Level points score variation between institutions, between cohorts within institutions, and between students within cohorts and within institutions,

2. a 'linear' model (4) with fixed terms for the three cohorts of students with 1993 as the base year and average GCSE points score as the best fitting explanatory attainment measure, and

3. a more complete prior attainment model (5) with fixed term higher order GCSE attainment measures

For the prior attainment model (5), the GCSE coefficients indicate the number of standard deviations the predicted total A-level points would increase by if the explanatory variables were increased by one standard deviation. For example, an increase in average GCSE points of one standard deviation - or just under 1 point - is associated with an increase in total A/AS Level points of 0.567 standard deviations - or about 5.5 points.

3.6. The coefficients of the GCSE measures in model 5 are positive (and statistically significant) confirming the nature of the non-linear relationship between the GCSE and A/AS Level measures. There is some evidence to show that the average progress (or value-added) has changed slightly over time: both year coefficients are negative. Average progress by the 1994 cohort was less than that of the 1993 cohort but progress in 1995 (whilst being less than in 1993) was slightly higher than in 1994. The correlation between the normalised GCE A-level scores and the total and average GCSE scores respectively are 0.63 and 0.71 compared with 0.62 and 0.70 for the un-normalised data: this suggests that the distributions of attainment are robust with respect to normality and the transformations do not have to be reworked separately for each subset of the data, which may vary between models.

Table 3.2. Joint prediction of A level score from total and average GCSE scores

	Prior attainment model (5)		Intercept model (3)		Linear Model (4)	
Variable	estimate	s.e.	estimate	s.e.	estimate	s.e.
Fixed:						
Intercept (α)	-0.132	0.007	-0.1469	0.0092	-0.027	0.0047
1994	-0.028	0.004			-0.010	0.0037
1995	-0.018	0.004			-0.011	0.0037
Total GCSE	0.138	0.002				
GCSE Average	0.567	0.002			0.648	0.0011
GCSE Average2	0.127	0.002				
GCSE Average3	0.0062	0.0004				
Spline term	0.020	0.005				
Random						
σ_v^2	0.04 (8%)	0.001	0.22 (22%)	0.0063	0.04 (8%)	0.0013
σ_u^2	0.01 (2%)	0.0003	0.01 (1%)	0.00045	0.01 (1%)	0.00033
σ_e^2	0.42 (90%)	0.0008	0.77 (77%)	0.0015	0.45 (90%)	0.00089
% total variation explained	53.5		0.00		50.4	
intra- year correlation: %	1.7		0.9		1.7	
intra- school correlation: %	8.2		22.1		8.1	

Note: The Level 3 variance is between institutions, level 2 is variance between cohorts (years) within institutions and level 1 is the residual or unexplained variance between students

.

3.7 Table 3.2 allows insight into the variation between students' A/AS Level scores before and after adjustment for their GCSE attainments. Most importantly, the use of multi-level models allows the information contained within the dataset's hierarchy (with students within cohort within institution) to be examined at each of the three levels: σ_v^2 refers to the between institution variation, σ_u^2 the within institutions between cohort component and σ_e^2 to the within institution within cohort between student variation.

3.8 Model 3 - the 'intercept' model' - shows that although three quarters (77 per cent) of the variation between students in their A/AS Level scores occurs between individuals within institution and cohort, over one fifth (22 per cent) occurs between institutions. Models 4 and 5 show the variation between students' A/AS Level scores once their GCSE achievements are taken into account: these provide measure of 'progress' or 'value-added' between GCSE and A/AS Level.

3.9. Model 4 indicates that 50 per cent of the total variation in students A/AS Level scores is explained by their average GCSE score. Conditioning on both average and total GCSE measures and adding polynomial terms as in Model 5 increases the amount of variation explained by about 3 percentage points, to 53.5 per cent. Both Model 4 and 5 indicate that, on average, some 90 per cent of the variation in value-added is between individuals within institution and cohort, with 8 per cent between institutions. The inclusion of student prior attainment reduces the variation in A/AS Level scores between institutions by 83 per cent (from 0.22 in Model 3 to 0.04 in Model 5 in table 3.2). We note that the between cohort component of the between institutions variation is relatively small in both Models 4 and 5: about two percent of the variation for Model 5. We shall elaborate on this point later in the paper. Finally, we see that the estimated between-institutions differences -

or 'residuals' - from Model 5 can be used as a measure to distinguish institutions which appear, given this particular model, to have students whose progress between GCSE and A/AS Level is, on average, statistically significantly high, or low.

4. Student Level Influences

4.1. The fixed part of Model 5 - primarily prior attainment measures - accounted for 53.5 per cent of the variation between students' A/AS Level scores. We now add to this model further terms - such as student gender, age, institution type attended on their GCE course and whether they changed institution between GCSE and GCE - to see whether the explanatory power of the fixed part of the model can be improved.

4.2. Model 6 in Table 4.1 shows the effect of including student gender. With the 'base' being the attainment of boys, we note that, on average, girls make less progress in total A/AS level performance between 16 and 18 than boys by about two thirds (.65) of a GCE A-level point. Furthermore, the interaction between prior attainment at GCSE and gender on progress is statistically significant. Whereas lower GCSE attaining girls make similar progress to boys, higher attaining girls make less progress; for example, for those gaining an average of grade A, girls make, on average, about 2 GCE A-level points less progress than boys. This broad conclusion cannot reflect more subtle connections between progress and gender: for example the differential effects on progress dependent on institution type attended or on subject choice at A/AS Level. The relationship between gender and progress is not linear, as is shown by the significance of the polynomial interaction terms in the fixed part of the model.

4.3. Model 7 in Table 4.1 suggests there is a slight negative relationship between progress made and the age of a student. In this context student age is an effect measured within a twelve month period being the maximum difference in age of students included in the dataset. Although this contrasts with results from previous studies which generally find that older students of compulsory school age made slightly more progress (for example Thomas and Mortimore,1994) - the effect is small and does not reduce the variation in the fixed part of the model nor the distribution of the variation between the levels of the hierarchy.

Table 4.1

Gender and age effects on student progress. [Higher order coefficients fitted as in table 3.5, but details omitted

	Model 6		Model 7	
Variable	estimate	s.e.	estimate	s.e.
Fixed:				
Intercept	-0.096	0.007009	-0.09	0.007
1994	-0.027	0.003	-0.028	0.0037
1995	-0.017	0.004	-0.018	0.0037
Girl	-0.067	0.003	-0.068	0.0026
Total points	0.14	0.002	0.14	0.002
Average points	0.61	0.003	0.61	0.003
Average *Girl	-0.067	0.003	-0.068	0.0024
Average2 *Girl	-0.016	0.002	-0.016	0.0019
Average3 *Girl	-0.001	0.0006	-0.001	0.00057
Age in Months			-0.0087	0.00027
Random				
σ_v^2	0.04(8%)	0.001194	0.04(8%)	0.0012
σ_u^2	0.01(2%)	0.0003169	0.01(2%)	0.00032
σ_e^2	0.41(90%)	0.0008297	0.41(90%)	0.00083
% total variation explained	54.2		54.2	
intra school correlation: %	8.1		8.1	

4.5. We also modelled the effect on student progress to A/AS Level of changing institution after GCSE. For simplicity, and to avoid large amounts of missing information on school movement for 1993 and 1994, we confined our analysis to the 1995 cohort. Table 4.2, Model 8, describes this analysis. We have included in the fixed part of the model a dummy variable for each student indicating whether the institution at which they took GCSE and A/AS Level exams was different . The coefficient is negative indicating that a student who changed institution made less progress. Although statistically significant, the effect is very small and, as for the effects of student age,, the proportion of variation explained by the fixed part of the model has hardly altered from that in Model 6 (Table 4.1).

Table 4.2

Impact of changing institution between GCSE and GCE A-level - 1995 cohort

	Model 8	
Variable	Estimate	s.e.
Fixed:		
Intercept	-0.09951	0.007169
GCSE Total	0.1454	0.002284
GCSE Average	0.6061	0.002761
GCSE Average2	0.1404	0.0019
GCSE Average3	0.0075	0.00047
Spline term	0.018	0.0049
Girl	-0.067	0.0026
Average * Girl	-0.067	0.0024
Average2 *Girl	-0.016	0.0019
Average3 *Girl	-0.0013	0.00058
Institution Change	-0.0087	0.00026

5. Institutional influences on Value Added

5.1 In England, there are a number of different types of institutions which provide 16-18 education leading to GCE A-levels. For this analysis we have sub-divided them into:

- LEA comprehensive schools
- LEA grammar schools (selective)
- LEA other maintained schools
- Grant Maintained (GM) grammar schools (selective)
- GM comprehensive schools
- Independent schools (Selective)
- Sixth Form Colleges
- Other FE colleges

Admissions policies vary across institution types and between institutions within the same type. Some institutions cater for single sex entry. Some select for ability. The socio-economic environment around a school can vary: both the mean attainment and the range of attainment can differ as can social class and the ratio of the different sexes. We are unable to control for student intake characteristics such as social class. Nevertheless it is of interest to know whether average progress changes according to institution type even though we recognise that any apparent differences will, in part, reflect the socio-economic profile of students attending particular types of schools (for example see impact of percentage with free school meals, Table 5.4, and the compositional effects Table 5.3).

5.2 Model 9 in Table 5.1 includes in the fixed part dummy variables for the 8 institutional types mentioned in paragraph 5.1 with sixth form college as the 'base' category. The coefficient for each dummy tells how much student progress differs for that particular institution type relative to sixth form colleges in terms of standard deviations of total GCE A-level points. Because there was insufficient data available on institution type in the dataset for candidates who sat GCE A-levels in 1993, the analysis covers the 1994 and 1995 cohorts with 1994 data as the 'base' period.

5.3. The coefficients show unexplained institutional effects, after allowing for prior attainment and gender (but not socio-economic or compositional factors). Those for grant maintained schools and for LEA grammar schools were not significantly different from zero, and therefore their students' progress are judged to be similar to that of students in sixth form colleges. Students in LEA comprehensive schools made slightly less progress than those in sixth form colleges - by about 3/4 of an A Level point - and students in FE colleges made approximately 2.5 points less progress. Students in selective independent schools made the most progress, averaging about 1.3 points more than those in sixth form colleges. Progress made by students in FE colleges is, however, likely to be understated when using total GCE A-level points as outcome as some of them will have taken a combination of GCE subjects and vocational qualifications which are not measured in this analysis.

Table 5.1

Value Added by Institution type 1994-1995

	Model 9	
Variable	Estimate	s.e.
Fixed:		
Intercept	-0.12	0.012
1995	0.011	0.0033
GCSE Total	0.13	0.0023
GCSE Average	0.59	0.0033
GCSE Average2	0.14	0.0023
GCSE Average3	0.008	0.00058
Spline term	0.0084	0.0055
Girl	-0.065	0.003
Average * Girl	-0.066	0.0029
Average2 *Girl	-0.015	0.0024
Average3 *Girl	-0.077	0.00058
Comprehensive	-0.078	0.016
Selective Grammar	0.03	0.026
Other Maintained	-0.11	0.029
Selective GM	0.028	0.24
Non Selective GM	-0.029	0.018
Selective Independent	0.13	0.017
FE	-0.25	0.019
Random:		
σ_v^2	0.027(6%)	0.001
σ_u^2	0.006(1%)	0.00037
σ_e^2	0.41(92%)	0.001
% total variation explained	55.1	
intra school correlation: %	6.1	

The variation explained by the introduction of the institutional type variables has increased the variation in A/AS Level performance explained by 2 percentage points, evidenced by the reduction in between institution variation to 6 per cent.

5.4 Models 10, 11, and 12 in Table 5.2 examine the relationship of entry policies to progress made by students in school sixth forms (i.e. excluding FE and sixth form colleges). These models cover the 1994 and 1995 cohorts but are otherwise identical to Model 5. For ease of interpretation the polynomial coefficients have been excluded from the table. Model 10 includes dummies for single sex entry schools (with mixed sex entry as base category): both single sex categories have statistically significant positive coefficients indicating that students in single sex schools make significantly more progress than those in mixed sex schools, in both cases by, on average, 1 GCE A-level point. Furthermore, boys schools make slightly more progress than girls schools (by a very small amount) which may be expected given that higher attaining boys make more progress than girls.

5.5. Model 11 examines admissions policies, where other maintained schools are the base category. Students in grammar schools make on average 2 A Level points more progress than those in comprehensive schools, who make 0.5 of an A Level point more progress on average than students in other maintained schools. This confirms the results of Model 9 in Table 5.1 where students at grammar schools made more progress .

5.6. When both types of entry policy (gender and selective/non-selective) are included together in the fixed part, in Model 12, a different story emerges. Progress by students in single sex schools are now no longer significantly different from students' progress in mixed schools: differential progress is dependent on whether students are in selective schools or not. The single sex school factor therefore seems to be accounted for by admissions policy: and most single sex schools are selective.

Table 5.2

Value-added according to school admission policy, 1994 and 1995 cohorts

	Model 10		Model 11		Model 12	
Variable	estimate	s.e.	estimate	s.e	estimate	s.e.
Intercept	-0.086	0.0069	-0.22	0.01	-0.22	0.01
1995	0.0085	0.0023	0.012	0.0034	0.012	0.003
GCSE Average	0.62	0.0032	0.89	0.0033	0.59	0.003
GCSE Total	0.15	0.0021	0.15	0.0024	0.15	0.002
Girl	-0.086	0.0033	-0.066	0.0032	-0.066	0.0032
Boys' School	0.11	0.0042			0.02	0.015
Girls' School	0.09	0.0041			0.02	0.013
Selective			0.24	0.011	0.22	0.013
Comprehensive			0.057	0.0097	0.056	0.0097

5.7. We now turn attention to the effects on progress in schools of two institutional-level attainment factors, or compositional effects. Students in FE and sixth form colleges were omitted from those analyses, as they were from Models 10 -12 in Table 5.2. The specification of Models 13 and 14 in Table 5.3 are otherwise as in Model 5 in Table 3.2. In Model 13 we look at how progress varies according to the range and average prior GCE attainments in schools and by the proportion of girls to see whether there are 'peer group' effects on progress. In Model 14 we attempt to judge whether sixth form size has a bearing on student progress.

5.8. The two attainment factors in Model 13 are the school average of students' average GCSE scores, and its standard deviation. Both coefficients are positive, indicating firstly that the higher the mean GCSE score in a sixth form the greater the average progress. This might be expected since higher average GCSE scores are obtained by students in grammar schools where students generally make more progress. Secondly, once student-based attainments and characteristics and the school-level mean average GCSE score have been taken into account, students in schools where the variation in their attainment around the school-level average score is higher also tend to make more progress.

5.9. In Model 14 we included a variable for the numbers of A/AS Level candidates within an institution. The coefficient of the variable is positive, indicating that those in larger A/AS Level sixth forms made more progress. However the coefficient is very small, indicating that the relative effect of sixth form size has limited impact on progress.

Table 5.3

Impact of compositional factors in schools

	Model 13		Model 14	
Variable	estimate	s.e.	estimate	s.e
Intercept	-0.12	0.021	-0.14	0.008
1994	-0.026	0.0037	-0.27	0.0037
1995	-0.016	0.003	-0.017	0.0037
GCSE Total	0.13	0.002	0.14	0.002
GCSE Average	0.61	0.003	0.61	0.0027
Girl	-0.066	0.0026	-0.67	0.0027
Average * Girl	-0.068	0.0024	-0.068	0.0029
Average2 *Girl	-0.016	0.0019	-0.016	0.0019
Average3 *Girl	-0.00096	0.00058	-0.001	0.00058
School Total GCSE Mean	0.1325	0.0068		
School Total GCSE SD	0.072	0.02		
Proportion of girls	-0.048	0.015		
No. of candidates			0.00024	0.00002

5.10. Finally, in Model 15 in Table 5.4, we look at the relationship between a measure of social class and student progress for students in the 1994 and 1995 cohorts combined. DfEE does not collect information on the socio-economic background of individual students, nor is any reliable institution level variable available from other sources such as the Census. However, schools do provide the department with information on the proportion of free school meals (FSM) taken up by students in the compulsory school age groups. Free school meals entitlement is income related and many studies have established strong statistical associations between this measure and social disadvantage.

5.11. Since this variable only applies to institutions which have students in this age group, FE and Sixth Form Colleges have been excluded from Model 15. Furthermore, because the FSM variable does not cover the age groups assessed in this analysis and since there is substantial student mobility between institutions post-GCSE, it can only at best be a crude indicator of the social class of the students in sixth forms. The variable - which has been standardised - has a negative coefficient indicating that the higher the proportion of students taking free school meals, the less the progress made by students in sixth forms.

Table 5.4

Value-added by free school meals take-up 1994-1995

	Model 15	
Variable	estimate	s.e.
Fixed:		
Intercept	-0.081	0.011
GCSE Total	0.16	0.0059
GCSE Average	0.57	0.0042
GCSE Average2	0.15	0.0036
GCSE Average3	0.01	0.00091
Spline term	0.05	0.0096
Girl		
Average * Girl	-0.053	0.005
Average2 *Girl	-0.049	0.003
Average3 *Girl	-0.009	0.001
Free School Meals	-0.066	0.0037
Random:		
σ_v^2	0.035(8%)	0.0013
σ_e^2	0.387(92%)	0.002
% total variation explained	56.5	
intra school correlation: %	8.3	

6. Stability of the model

6.1. In this section we explore the sensitivity of progress models over time and to the exclusion of students attempting only 1 A Level.. The Gray et al (1995) analysis was based on students who had attempted two or more GCE A-level subjects. In our analysis so far we have considered progress for students who attempted 1 or more GCE A/AS Level. Accordingly, we re-ran Model 6 excluding the 4478 students in our 3 year dataset who attempted only 1 A/AS Level. Table 6.1 shows the results of this analysis. As expected, on average, there was very little change in the values or relative strengths of the coefficients or in the distribution of the variation of progress between the levels of the hierarchy.

Table 6.1

Stability of model estimates: exclusion of students attempting 1 A/AS Level

	1 or more A-level Attempts		2 or more A-Level Attempts	
	Model 6		Model 16	
Variable	estimate	s.e.	estimate	s.e.
Fixed:				
Intercept	-0.096	0.007	-0.097	0.007
1995	-0.027	0.003	-0.014	0.0037
1994	-0.017	0.004	-0.025	0.0037
GCSE Total	0.14	0.002	0.13	0.002
GCSE Average	0.61	0.003	0.61	0.0027
GCSE Average2	0.14	0.0019	0.14	0.0019
GCSE Average3	0.0074	0.00047	0.0073	0.00048
Spline term	0.02	0.0049	0.015	0.0049
Girl	-0.067	0.003	-0.067	0.0026
Average * Girl	-0.067	0.003	-0.68	0.0024
Average2 *Girl	-0.016	0.002	-0.017	0.0019
Average3 *Girl	-0.001	0.0006	-0.0013	0.0006
Random:				
σ_v^2	0.037 (8%)	0.001194	0.036 (8%)	0.0012
σ_u^2	0.0083 (7%)	0.0003169	0.0082 (7%)	0.00032
σ_e^2	0.41 (90%)	0.0008297	0.41 (90%)	0.00083
% total variation explained	54.2		54.3	
intra class correlation: %	8.1		7.9	

6.2 The Gray et al (1995) analysis - and to a more limited extent the previous DfEE technical work - looked at the stability of progress model coefficients over time by fitting separate, but identical, models for each year. Models 17, 18, and 19 in Table 6.2 (based on Model 5) show that the coefficients and their standard errors to be similar across the three years, 1993, 1994, and 1995. There is some evidence that the relationship between average GCSE points score and total GCSE points score changed over the 3 years with the association between A/AS Level scores and total GCSE points increasing, and those on average GCSE decreasing. In 1995 the spline function became not significantly different from zero. In part, this might reflect the move by students over the period to attempt more GCSE examinations.

Table 6.2. Stability of model coefficients over time

	1993		1994		1995	
	Model 17		Model 18		Model 19	
Variable	estimate	s.e.	estimate	s.e.	estimate	s.e.
Fixed:						
Intercept	-0.064	0.01	-0.11	0.01	-0.13	0.01
GCSE Total	0.112	0.0036	0.13	0.0036	0.17	0.0032
GCSE Average	0.64	0.0046	0.6	0.0047	0.58	0.0047
GCSE Average2	0.13	0.0032	0.14	0.003	0.15	0.0033
GCSE Average3	0.0054	0.00077	0.0062	0.00085	0.0095	0.0008
Spline term	0.036	0.0086	0.021	0.0084	-0.0022	0.0083
Girl	-0.07	0.0046	-0.072	0.0046	-0.064	0.0044
Average * Girl	-0.07	0.004	-0.07	0.0046	-0.06	0.0042
Average2 *Girl	-0.016	0.0033	-0.016	0.0033	-0.014	0.0034
Average3 *Girl	-0.000046	0.0009	-0.0018	0.001	-0.001	0.001

6.3 Table 6.3 shows the correlations between the institutional residuals for each year derived from separate models for the three yearly datasets (i.e. models 17-19) and compares them with those obtained from a single model incorporating the combined three year dataset. The correlations between years for the yearly datasets are quite low at 0.63 and 0.68 for the 1993/94 and 1994/95 correlations respectively indicating a lack of stability between years. These correlations are very similar to that produced by Gray et al. (1995) for 1993/94. The correlations produced using the combined three year dataset are however much higher at 0.9 and 0.94 for 1993/94 and 1994/95 respectively, indicating a greater degree of stability when a fixed part averaged over the three year period is used. This demonstrates that estimates of value added are more stable when institutional effects are modelled in a single multi-year dataset which incorporates the time level as a random variable.

Table 6.3

Correlations of institutional effects over time using three one year datasets

	Three Yearly datasets		Combined three year Dataset	
	1993	1994	1993	1994
1994	0.63		0.90 (0.822)	
1995	0.57	0.68	0.84 (0.754)	0.94 (0.869)

6.4 The true correlations between sets of residuals are lower than those calculated from the 'shrunken' institutional effects because the latter reduce the impact of institutions with small numbers of students. The multilevel model software (MLM) does estimate the true correlations between years, given in the brackets in Table 6.3. Thus, for example, the true correlation between the 1993 and 1994 residuals is 0.82 compared to the shrunken estimate of 0.90.

6.5. Gray et al (1995) also investigated how the relationship between GCE A-level and GCSE changed when the outcome variable changed from total A-level points to average A-level points. They found that the relative values of the coefficients were similar regardless of which outcome measure was used. They also investigated the influence of General Studies and found that the inclusion of this subject in the GCE A-level points score would cause the effectiveness measures for individual institutions to vary although only for a small number of them. In general most institutions who did well using one outcome measure did well using the other and vice versa. Although we have not carried out these sensitivity analyses, our 3 year dataset would allow such work to be done.

7. School and college improvement

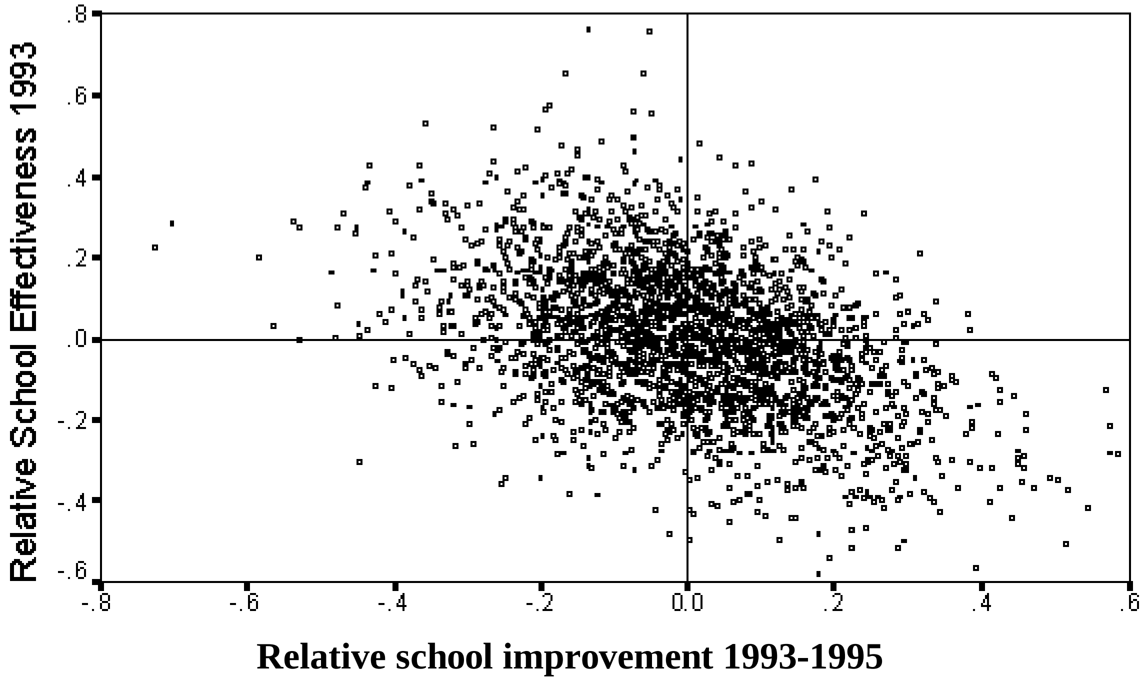
7.1. One of the outputs from the multilevel models we have analysed are institutional-level 'residuals' which can, together with those residuals' confidence intervals, provide statistical estimates of the average progress made by students in an institution relative to that of other institutions. This 'hard' information could be used by schools and colleges to assist their improvement and effectiveness development planning. Using our dataset we have investigated progress:-

- on average by students within individual institutions , and
- over time, to see if there is evidence that student progress has varied
- according to the assumptions in model structure
- by groups of students categorised by levels of GCSE attainment to determine whether students within institutions make differential progress (discussed in Section 8)

7.2 We have carried out an analysis of institutional effectiveness and improvement similar to that conducted by Gray, Goldstein and Jesson (1996), who investigated school improvement over a five year period and how it varied according to school effectiveness at the beginning of the period. 'Institutional improvement' can be measured by comparing schools/colleges residuals (or 'effect') over the time-span of the dataset: in our case the difference between institutional effectiveness in 1993 and 1995 provides an estimate of improvement relative to other institutions. Multilevel Model 6 (Table 4.1) has been used to estimate an overall effect for each institution for each of the three years.

7.3. Figure 7.1 plots institutional improvement between 1993 and 1995 against the level of effectiveness in 1993 for a 10 per cent random sample of institutions from the dataset. All measures are relative values, so that an institution with a zero effect on either axis is the average institution. Figure 5.1 can be sub-divided into quadrants to ease interpretation. The bottom left quadrant contains institutions which had a less than the average institutional effect in 1993 and below average improvement between 1993 and 1995: approximately 17% of institutions fell into this category. The bottom right hand quadrant contains those which had below average effectiveness in 1993 but above average improvement, approximately 33% of institutions. Similarly, the top two quadrants contain institutions showing above average effectiveness in 1993 with below, and above, levels of improvement over the period: approximately 33% and 17% of institutions respectively fell into these two groups.

Figure 7.1
School and College improvement over three years versus effectiveness in 1993.



7.4. Both the effectiveness and improvement estimates have associated uncertainty intervals which are estimated by the multilevel model software. By constructing confidence intervals around each estimate, we can determine whether that institution is statistically significantly different from the average (i.e. if the confidence interval around the residual does not contain zero). The estimated effects are based on the progress of all students in an institution and, for practical purposes, reflect most closely the institutional effectiveness for average students. We discuss in Section 8 the question of differential institutional effects and the different rates of progress that take place at the extremes of GCSE attainment.

7.5. Institutions whose effects in 1993 had confidence intervals containing zero are deemed not significantly different from average. Positive effects whose uncertainty bands do not contain zero are deemed above average, while negative effects are taken as below average. The improvement groups were constructed in the same manner, using the difference between effectiveness in 1993 and 1995. Confidence intervals are conservative as covariances between groups have been assumed to be zero. For graphical methods of comparing school residuals see Goldstein and Healy 1995.

7.6. Cross-classifying these groupings we consider in Table 7.1 the nine categories covering the possible combinations of effectiveness and improvement.

Table 7.1

Changes in Value Added by effectiveness in 1993, by institution type in 1995: percentages

1993 Effect	Improvement 1993-1995	County	FE	Indep.	GM	VA	VC	Total
Average	No Change	66	45	56	64	62	61	59
Below Average	No Change	11	25	2	8	16	4	10
Above Average	No Change	8	11	31	14	11	10	15
Average	Below Average	4	9	2	2	3	4	4
Below Average	Below Average		2					0.3
Above Average	Below Average	3	5	4	4	1	5	4
Average	Above Average	4	2	3	6	4	8	4
Below Average	Above Average	4	1	1	4	3	6	3
Above Average	Above Average			1			3	0.3
	Total Percent	100	100	100	100	100	100	100
	Total Number	1020	387	572	397	196	80	2604[3]
Note: County schools are LEA schools and exclude voluntary aided (VA) and voluntary controlled (VC) shown separately								

The categories have analysed these by institution type. City technology colleges and special schools have however been omitted from the table because of their small numbers .

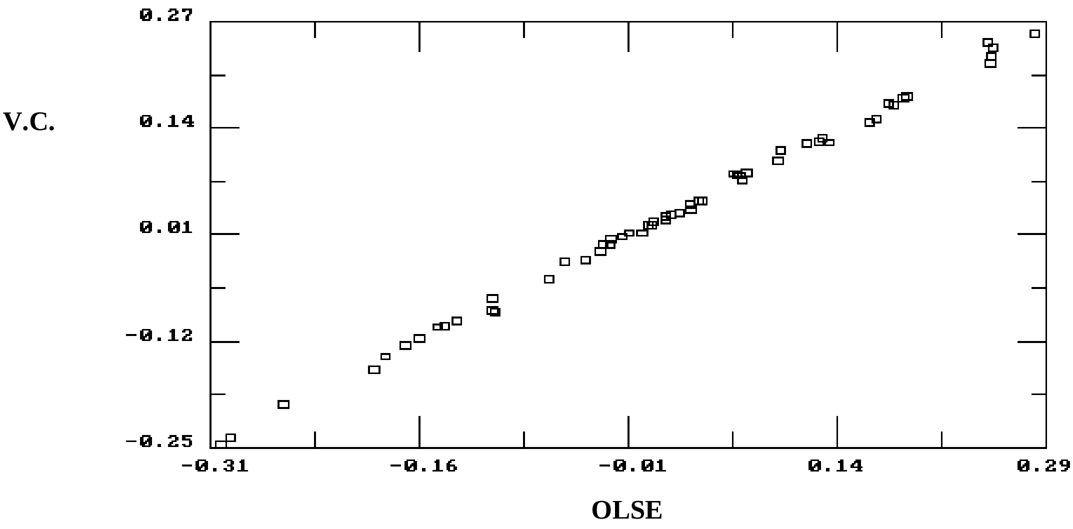
7.7. Most institutions (59%) were in the average effectiveness group in 1993 and did not show significant improvement in their effectiveness between 1993 and 1995. A further 25%, that were above or below average in 1993 did not significantly change relative to the average over the period. Accordingly 16% showed significant change in levels of effectiveness with equal proportions showing above and below levels of improvement. In 1993 some 33% of institutions were significantly more or less effective than average (compared with 35% in 1995) and of these 13% had below average levels of effectiveness, and 19% were judged to have been of above average effectiveness.

7.8. Proportionately more voluntary controlled schools (VC) - of which there were 80 in the dataset - improved their effectiveness relative to other institutions: 17% of these institutions significantly improving more than the average. FE colleges had the highest proportion showing below average levels of improvement, with 16% of these institutions in this category. Note however that the indicators in Table 7.1 are of the improvement for average GCSE attainers: institutions which have proportionately more (or less) low (or high) GCSE attainers a measure of progress of the average student may not be relevant. County maintained schools were slightly more likely to be of average effectiveness in 1993 than other state schools and to show slightly less propensity to have improved effectiveness than grant maintained schools. Over one third (36%) of selective independent schools were of above average effectiveness in 1993 and were less likely to evidence either above or below average levels of improvement over the three years to 1995.

7.9 The measured institutional effects could depend upon the type of model chosen for the analysis. To test this we investigated whether the results from a linear multilevel Model 4 gave different answers from an equivalent OLS regression specification. The OLS model fitted the average GCSE score using parallel straight lines, one for each year, averaging the raw residuals for each institution over the three year period to compile ‘institutional effects’. This OLS analysis has been widely used and is, for example, essentially the specification adopted by ALIS. Figure 7.2 plots the school residuals for these OLS and MLM models. The relationship between the two sets of residuals is highly linear, and we conclude that the for the purpose of calculating overall institutional effects which reflect the performance of average students, the choice of OLS or MLM specification may not be critical. But, as Section 9 demonstrates, the confidence intervals associated with OLS estimates are biased and less efficient (i.e. larger) than those of the MLM estimates and will provide less discrimination between institutions.

[3] 2604 is the number of schools which existed over the three years and not the sum of the column totals. City technology colleges and specials schools are not included in the columns.

Figure 7.2 Variance components analysis fitting straight line vs. OLSE



8. Residual analysis by GCSE group

8.1. We now investigate whether there is any evidence to suggest that some institutions are differentially effective. For the purposes of this analysis an institution is said to be differentially effective if there is evidence that a student's progress to A/AS Level varies according to their average or total GCSE attainments. For example, in some institutions, students with higher initial attainments may make more progress, compared with lower attaining students in that institution, than they would do in other institutions. Other student characteristics, such as ethnicity may also be examined, but this is beyond the scope of the present paper. Goldstein and Thomas (1996) have found evidence of institutional differential effectiveness, both by gender and ability. Gray et al. (1995) found that 11% of institutions were differentially effective with regard to ability, and of these 56% tended to do better for more able students and 44% with less able students. We shall compare the relative effectiveness of institutions for different levels of student prior attainment and use :-

- Model 6 directly for the polynomial MLM specification and equivalent OLS model, and
- MLM and OLS models linear in average GCSE score

8.2. So far all the MLM (and OLS) models have not separately identified students according to GCSE attainment: we now wish to look at the variation between students within institutions - (2e - has remained undivided. For ease of interpretation, we partitioned students into five groups by their normalised GCSE scores. Thus:-

- lowest decile (0%-10%)
- next two deciles (10% - 30%)
- next four deciles (30% - 70%)
- next two deciles (70% - 90%)
- the top decile (90%-100%)

and then added the appropriate dummy for each individual student to the appropriate random part of the model, (2e . In Table 8.1 we have compared the correlations between institutional residuals for each of the 5 student GCSE attainment bands separately for OLS and MLM specifications based on a linear model.

Table 8.1

Correlations across GCSE groups for estimated residuals using OLS and MLM analyses. Institutions having at least 1 student in each GCSE group and at least 300 students over the three years. 364 institutions. Linear model.

OLS					
	10-30%	30-70%	70-90%	>90%	Mean (S.D.)
<10%	0.64	0.54	0.46	0.22	0.21 (0.32)
10-30%		0.79	0.63	0.31	0.07 (0.20)
30-70%			0.82	0.39	-0.03 (0.20)
70-90%				0.43	0.00 (0.22)
>90%					0.14 (0.38)
Multilevel					
	10-30%	30-70%	70-90%	>90%	Mean (S.D.)
<10%	0.93	0.71	0.58	0.49	0.37 (0.25)
10-30%		0.87	0.73	0.52	0.20 (0.19)
30-70%			0.96	0.68	0.03 (0.18)
70-90%				0.80	-0.02 (0.17)
>90%					0.04 (0.15)

8.3. We see that in every case the OLS model indicates a lower (or downwardly biased) correlations between GCSE bands than the MLM specification from which we would have concluded that institutions are more differentially effective than is the position were the more efficient MLM analysis to be used. The correlations between the residuals of the OLS and MLM models for each of the 5 groups are 0.84, 0.97, 0.98, 0.90, and 0.70. This shows that the correlations are lower for the extreme GCSE bands than for the students closer to average GCSE attainment, suggesting the two models make quite different estimates of the student effects at the extremes. We also see that for the linear models the mean of the residuals for each group is not zero for the lowest group in contrast to those of the polynomial-spline models shown in Table 8.2. This is because the linear model under and over estimates the relationship between the GCSE and A/AS Level indicators at the extremes of GCSE attainment. It is therefore important to fit in the fixed part of the model a function which properly graduates the relationships, and in general a straight line will not be adequate.

8.4. Table 8.2 presents correlations for the same GCSE groups in Table 8.1 but as output from the polynomial model. The MLM correlations between GCSE groups are all higher than for the linear specification whereas the OLS correlations for both model structures are not uniformly so. But the mean and standard error of the residuals for each of the 5 GCSE groups under both OLS and MLM specifications are now nearly zero which supports the use of a polynomial function to describe the GCSE and A/AS Level relationship. The correlations between OLS and MLM models for each of the 5 groups are 0.74, 0.95, 0.99, 0.90, 0.67. We note that these correlations are little different from the equivalent linear associations, but are lower for the bottom and top decile of students. These conclusions re-inforce both the importance of using MLM analyses (which uses all the information on attainment relationships contained in the dataset) in drawing conclusions on differential effectiveness from the institutional residuals, and the need to accurately describe the fundamental relationship between GCSE and A/AS Level measures across the whole attainment range.

Table 8.2

Correlations across GCSE groups for estimated residuals using OLS and MLM analyses. Institutions having at least 1 student in each GCSE group and at least 300 students in the institution over the three years. 364 institutions. Polynomial-spline model.

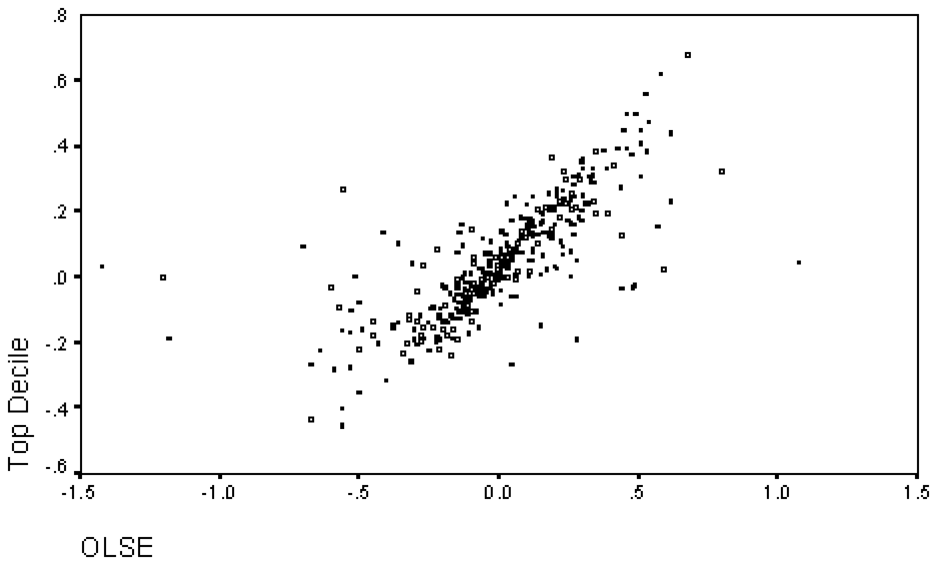
OLS					
	10-30%	30-70%	70-90%	>90%	Mean (S.D.)
<10%	0.57	0.47	0.41	0.28	0.00 (0.29)
10-30%		0.78	0.62	0.43	-0.01 (0.18)
30-70%			0.80	0.50	0.05 (0.18)
70-90%				0.52	0.06 (0.22)
>90%					0.08 (0.33)

Multilevel					
	10-30%	30-70%	70-90%	>90%	Mean (S.D.)
<10%	0.96	0.83	0.73	0.61	0.05 (0.18)
10-30%		0.93	0.82	0.71	0.05 (0.15)
30-70%			0.96	0.87	0.07 (0.16)
70-90%				0.96	0.07 (0.16)
>90%					0.06 (0.15)

8.5. Figure 8.2 highlights the difference between the residuals produced by the OLS and MLM analyses from the model specification underlying Table 8.2. Although the relationship between the sets of residuals is quite strong there are a substantial minority of institutions for which the OLS residual is very different from its MLM counterpart. In these cases the use of OLS estimates might well lead to the drawing of misleading conclusions about institutional effectiveness, for the reasons discussed above.

Figure 8.2

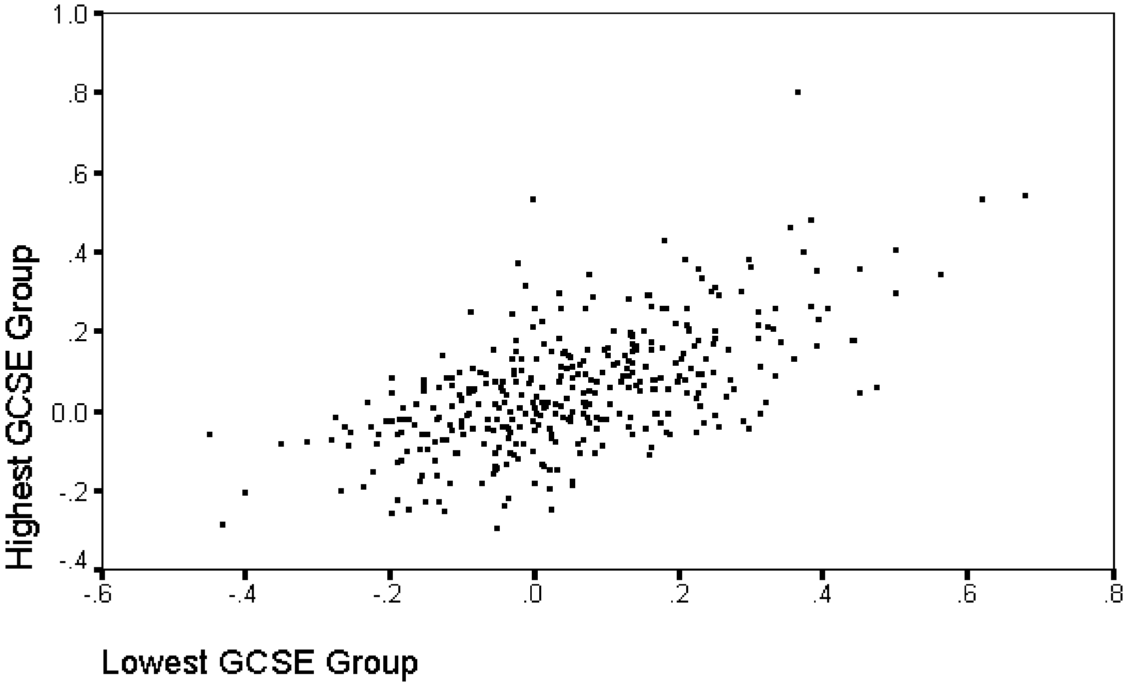
Students in the top decile: comparison of multilevel and OLS residuals from polynominal spline models underpinning Table 8.2



8.6. Figure 8.3 shows a scatterplot of the highest and lowest GCSE groups residuals for the 364 institutions which result from the MLM model of Table 8.2 (correlation 0.61). This confirms graphically that the data suggests differential institutional effectiveness. Although there is a distinct association between the progress that disparate GCSE groups of students make across institutions there is again solid evidence to show that in general differential effectiveness judged against prior attainment is not consistent across institutions; the relationship between progress made by the brightest and the weakest students within the same institution varies a great deal.

Figure 8.3

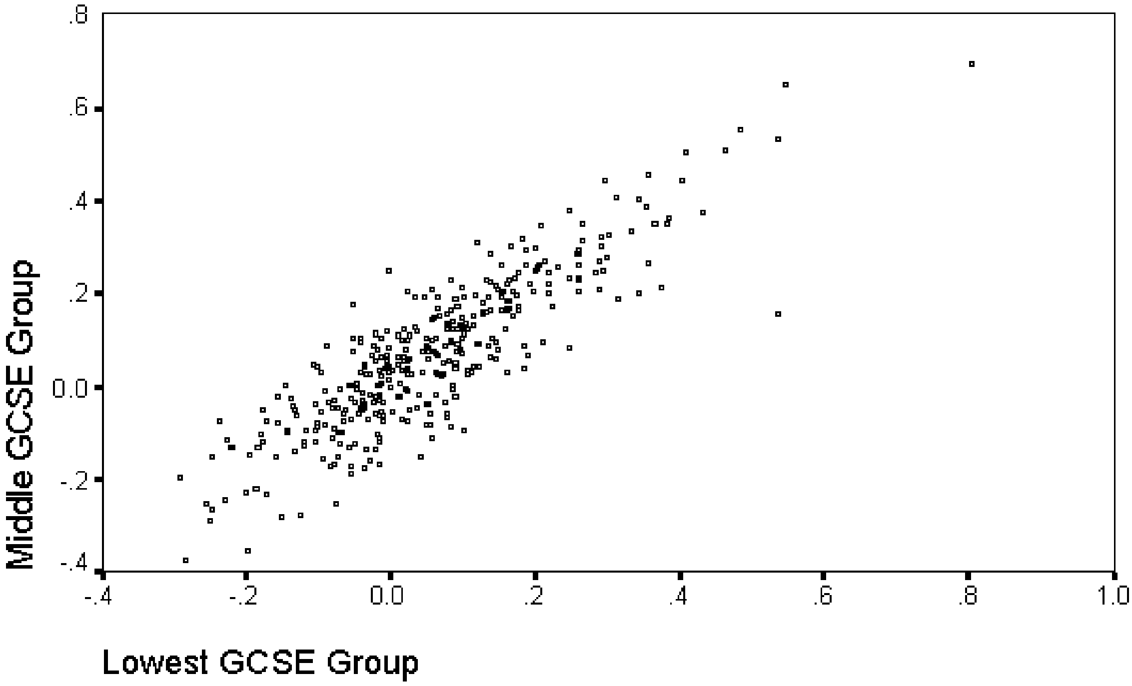
Students in the top and bottom deciles: comparison of residuals from MLM polynomial - spline model of Table 8.2.



8.7. Figure 8.4 shows that institutional differential effectiveness is not limited to the very extremes of student prior attainment. This graph shows the relationship between the progress made by students in the 30 - 70 per cent GCSE attainment range when compared with that of the lowest 10 per cent of students. Interestingly, there are a good number of institutions where there are apparently significant differences between the progress of both groups of students (correlations 0.83).

Figure 8.4

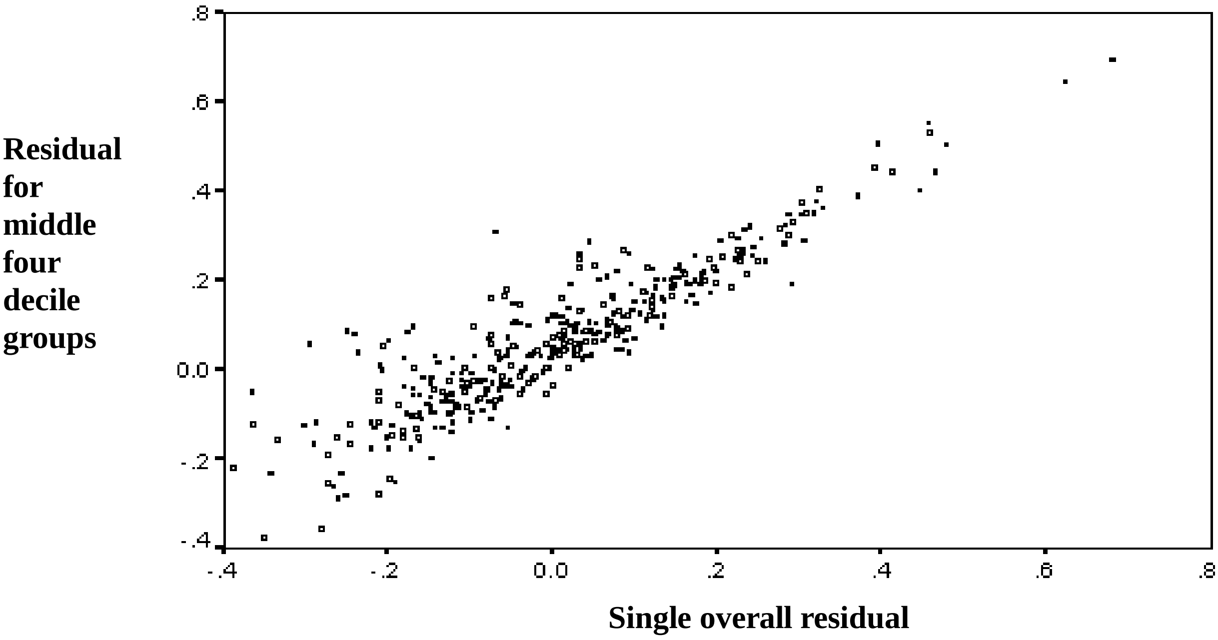
Students in the middle and lowest GCSE attainment bands: comparison of residuals from MLM model of Table 8.2



8.8. A single, overall, measure of institutional effectiveness reflects the progress of 'average' students. Whilst this may provide an adequate picture for some management purposes for institutions where the average student is typical, it yields little information about the distribution of progress across the ability range and hence is of limited value to the institution in its efforts to focus improvement strategies on specific groups of students. Figures 8.3 and 8.4 have illustrated how the GCSE/A Level dataset and the use of MLM models give powerful evidence of the existence of statistically significant differences in student progress over the GCSE achievement spectrum. As a final supporting indication of differential effectiveness, Figure 8.5 graphs residuals for the progress of students in the middle four deciles of the GCSE attainment range against residuals for all students. The relationship between the two sets of residuals is fairly strong but there is variation amongst institutions whose student progress is smallest. This information could provide the start for further diagnostic statistical examination.

Figure 8.5

Residuals for the middle four GCSE deciles against single overall residual for institutions in Table 8.2



9. Confidence Intervals

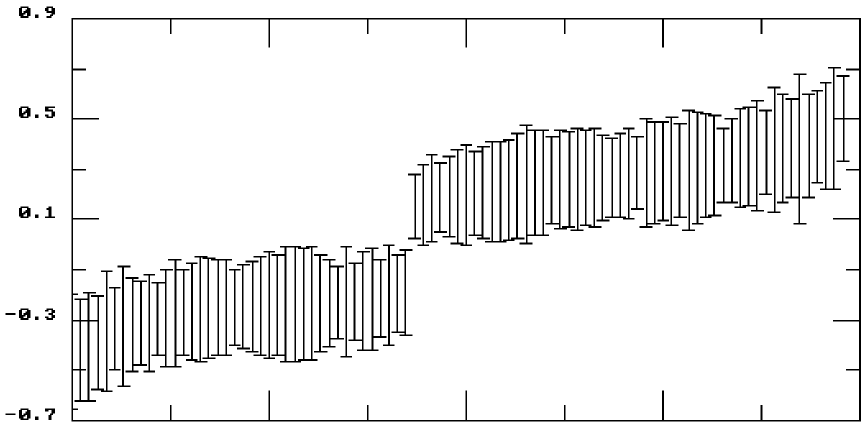
9.1. Each residual - an estimate of institutional effectiveness - whether overall or for individual prior attainment groups, has a standard error and associated confidence interval. This enables us to determine whether an institution's effectiveness is significantly better or worse than average (that is, the interval does not contain zero).

9.2. We constructed 95% confidence intervals for the multilevel polynomial-spline model underpinning Table 8.2. In Figure 9.1 we display the ordered residuals for the highest GCSE attaining group of students showing only those which do not overlap the mean residual (approximately zero). The figure shows that, of the 364 institutions relevant to the model, about 11% were judged to have students whose progress was significantly worse than the average, and 14% had students whose progress was significantly above average levels. Put alternatively, for three quarters of institutions the progress to A/AS Level of the highest attaining GCSE students was found to be not statistically distinguishable using standard confidence intervals.

Figure 9.1

Residuals of the >90% GCSE group with 95% confidence intervals which do not overlap the mean residual of that group: Institutions in Table 8.2

Residual

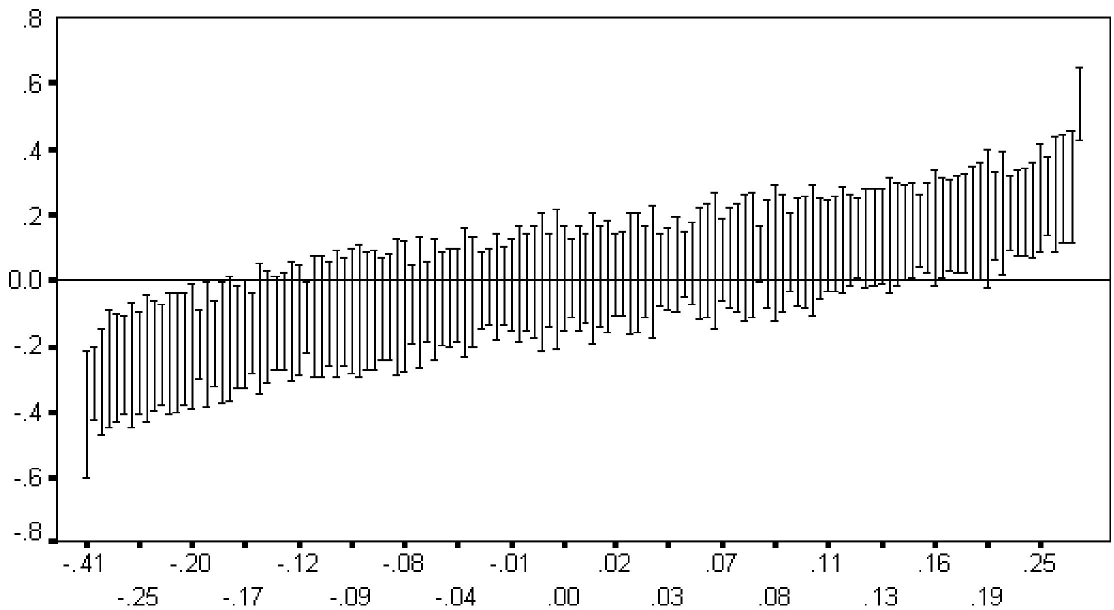


9.3. Figure 9.2 shows, again for Table 8.2 institutions, confidence intervals for the residuals of the difference between the progress of the highest and lowest GCSE attaining groups. For ease of presentation a one third random sample of the residuals is displayed. Only those institutions whose confidence intervals do not overlap zero can be said to be differentially effective with respect to the highest and lowest GCSE attainers. In approximately 19% of institutions the difference between these students progress was significantly less than average (i.e. they were less differentially effective than the 'average' institution), and for 15% of institutions the difference in student progress was significantly higher than the average level.

Figure 9.2

95% confidence intervals for the difference between the highest and lowest GCSE groups: 33% random sample of institutions in Table 8.2.

Residual



10. A theoretical comparison of multilevel and OLS residual estimates

The reason for the above differences can be seen by considering the following simple 2-level model involving just a straight line prediction of A - level (y) from GCSE (x).

$$y_{ij} = \alpha + \beta x_{ij} + u_j + e_{ij} \qquad (2)$$

The OLS estimates of the fixed part coefficients α, β are consistent and will be close to the multilevel estimates for a large sample of students. The OLS estimates for each school are

$$\tilde{y}_j = \sum_i \tilde{y}_{ij} / n_j$$
$$\tilde{y}_{ij} = (y_{ij} - \hat{\alpha} - \hat{\beta} x_{ij}) \qquad (3)$$

so that these are estimating not u_j but

$$u_j + \sum_i e_{ij} / n_j \qquad (4)$$

which is consistent as the number of students increases for a given school, but the (unconditional) between school variance is given by

$$\sigma_u^2 + \sigma_e^2 / n_j$$

so that, rather than being 'shrunken' they are 'stretched' with too large a variance. By contrast the variance of the multilevel between school estimates is given by

$$n_j \sigma_u^4 (\sigma_e^2 + n_j \sigma_u^2)^{-1}$$

and the ratio of this to the OLS variance is just the square of the shrinkage factor.

10.2 For small numbers of students the size of the OLS variance is determined by the size of the student level variance and can become very large. This is an undesirable feature. The multilevel shrunken estimates, however, have the property that for small sample sizes the estimates become close to the overall fixed part mean and the variance decreases. This seems desirable since where ther*e is little* information for a school (few students) the estimates become 'conservative'. By the same token, the estimated correlations between OLS estimates as in Tables 4 and 5 may be considerably attenuated. We also note that the conditional variance of the OLS estimator, which we use to construct confidence intervals, can be considerably larger than the multilevel one. For a variance components model the multilevel estimator and its variance (about the true value u_j) is given by

$$\hat{u}_j = n_j \sigma_u^2 (\sigma_e^2 + n_j \sigma_u^2)^{-1}$$
$$\mathrm{var}(\hat{u}_j) = \sigma_u^2 \sigma_e^2 (\sigma_e^2 + n_j \sigma_u^2)^{-1} \qquad (5)$$

The corresponding variance for the OLS estimator (4), however, is

$$\sigma_e^2 / n_j$$

Thus the ratio of the multilevel to the OLS variance is simply the shrinkage factor. Table 6 shows this for different sample sizes assuming a typical intra-unit correlation of 10%.

To obtain a consistent estimate of σ_e^2 from the OLS analysis we can use the 'pooled' estimator

$$\sum_j \sum_i (\tilde{y}_{ij} - \tilde{y}_j)^2 / (N - J)$$

where N is the total number of level 1 units and J is the number of level 2 units.

10.3 Table 10.1 shows the relationships between the MLM and OLS variances for institutions of various sizes. Most of the models created for the analyses in this paper have indicated an intra-institution correlation of about 10%. This is approximately the figure which other similar studies have found and we have used it in the table for convenience.

Table 10.1

Ratio of MLM and OLS residual (conditional) variances for different institutional year-group sizes and with intra-institution correlation of 10%

Unit size	MLM/OLS variance (and standard deviation) ratio
10	0.53 (0.73)
25	0.74 (0.86)
50	0.85 (0.92)
100	0.92 (0.96)
200	0.96 (0.98)

10.4 The Table shows that OLS estimators are less efficient than MLM estimates. They provide poorer discrimination between institutions unless the unit sample (or population) size is relatively large. This is clearly important when comparing institutions since it shows that the OLS estimator can provide a rather poorer discrimination, that is less efficient, unless the unit sample size is large. In terms of interpreting the model random parameters, that is the variances and covariances (correlations), the OLS approach is inadequate and attempts to do so will lead to biases.

10.5 In summary, these analyses have established the importance, in value added analyses, of properly modelling the relationships in the data to avoid bias, of carrying out efficient multilevel analyses and of studying differential effectiveness as a means of providing important information to institutions which cannot be captured by a single residual estimate. Finally, in a more general sense, it is the model parameters themselves which are of primary interest for understanding the factors associated with institutional differences and for this a multilevel approach is essential.

11. Conclusions and summary

11.1 This study carried out an analysis of value-added (or progress) between GCSE and A/AS Level of all students (over 500,000) reaching the age of 18 in schools and colleges in England for the three academic years 1993 to 1995. We studied the data using Multilevel Models (the simplest case being variance components models) with, in most cases three levels of hierarchy: students within year groups (cohorts) within institutions. We demonstrate that this method is theoretically more efficient than OLS regression, and show with practical examples from a large and comprehensive dataset that application of this method provides vital information about the relative performance of institutions and that of students of varying GCSE abilities within them.

11.2 We modelled the overall relationships between prior attainment (or input) measures in the form of individual students' average GCSE and total points score and an outcome measure expressed as total A-level points score. We fitted a polynomial and spline function model in total and average GCSE which was found to provide a significantly better fit for the extremes of GCSE attainment than did a linear model. This gave a superior reflection of the different value-added relationships across the student attainment distribution. In most of the models we defined, just over half of the variation in student A/AS Level attainment could be statistically explained by the measures of GCSE achievement: these prior attainment measures were by far the most statistically significant of the variables that we included in any of the models.

11.3. We included gender in the analysis. We found that girls, on average, progressed less than boys, but the relationship varied according to prior attainment. For weaker GCSE students girls progressed at a very similar rate to boys, but higher ability girls progressed at a lower rate than boys, up to about 2 GCE A-level points less on average for the very brightest students.

11.4. We added the institution type at which students took their A/AS Levels to the fixed part of the model and found that, on average, students at selective independent schools progressed more than at other institutions, and that progress was least at FE colleges. Changing institution between GCSE and A/AS Level had very small effects on progress. Single sex schooling was found to have no significant effect on progress once the admissions policy of institutions was included: students in selective schools made more progress. An alternative analysis including average student GCSE scores and their dispersion, but excluding direct estimators for types of institution showed that students in institutions where the average level of total GCSE points score was higher were found to make more progress than average, as did students in institutions where there was greater variation around the initial score. As expected , the higher the percentage of take-up of free school meals in schools, the lower the value-added between GCSE and GCE A/AS Level.

11.5. We investigated the differences in student progress for each of the three cohorts in the dataset to measure the stability of the institutional effectiveness over time. We found a three-year dataset produced more stable estimates of effectiveness than separate single year datasets. The actual correlation coefficients between the residuals of the 1993 and 1994 cohorts and between the 1994 and 1995 cohorts derived from the three year and one year datasets were 0.82 and 0.87 respectively. These relationships are of the same order as the correlations from year to year in institutions rankings on A/AS Level scores as published in the DfEE Performance Tables. We also examined whether there was evidence that overall measure (i.e. for the average student) of the effectiveness of each institution was changing over time. We compared their relative effectiveness in 1993 with that in 1995. About 33% of institutions were more, or less, effective in 1993 compared with 35% in 1995. Almost 60% of institutions were of 'average' effectiveness in 1993 and whose relative performance did not change over the 3 year period. Roughly 10% of institutions had below average and 15% above average levels of effectiveness in 1993 and did not significantly improve their relative performance by 1995. Some 16% of institutions did evidence changes in their relative effectiveness over the period of which 7% showed significant improvement.

11.6. We turned our attention from examination of differential effectiveness between institutions for the average student to focus on evidence for different rates of progress of students of differing GCSE attainments. We showed that students of the highest GCSE achievements made significantly greater or less progress than the average student in approximately one quarter of institutions, and that difference in progress between the highest and lowest GCSE attainers was significant in one third of institutions (with more than 300 students in each group). Thus - for these institutions at the least - an estimate only of their effectiveness for the average student would mask important information about specific student groups that might require closer management attention.

11.7. The existence of differential effects underlines the importance of modelling using multilevel methods and controlling carefully for prior attainment. The recommended polynomial spline model gave estimated institutional effects that differed from those provided by a simple linear model for students of above or below average ability. For these results to be correctly interpreted, unbiased estimates of confidence intervals about the measured institutional effects are required. They can be obtained only by taking account of the hierarchical nature of the data. Multilevel models allow for this, but OLS does not, and will give biased estimates of the confidence intervals.

11.8. Our analyses show the investigative power that resides within a large and well-known dataset when examined by efficient statistical models. These analyses - and others, such as by GCSE and A/AS Level subjects attempted - give clues about institutional performances which are statistically different from those of other institutions. They are only diagnostic tools and cannot give guidance as to the methods required to improve performance. However, they have a distinct place in helping to highlight aspects of performance that may be worthy of closer scrutiny, and thus have a role in institutional improvement.

11.9 Further work on the datasets would be desirable, not only to update the findings on receipt of later years' data, but also to study further the interactions between the factors involved, to further model the random coefficients and to allow for the possible effects of subject choice at GCE A level on student progress.

12. References

Aitkin, M. and N. Longford, 1986. ‘Statistical modelling issues in school effectiveness studies (with discussion)’, Journal of the Royal Statistical Society, Series A., Vol. 149, pp1-42.

Fitz-Gibbon, C.T., 1995. ‘The Value Added N*ational Project: preliminary report’, London: SCAA.*

Fitz-Gibbon, C.T., 1996. ‘Issues to be *considered in the design of a* n[at]ional value added system’, Value Added National Project: General Report, London: SCAA.

Goldstein, H. 198*7. Multilevel Models in Social and Educational Rese*arch. London: Arnold.

Goldstein, H. 1995. Multilevel Statistical Models, 2nd edn. London: Arnold.

Goldstein, H. and M. Healy, 1995. ‘The graphical presenta*tion of a set of means’, J*ournal of the Royal Statistical Society, Series A. Vol. 158.

Goldstein, H., J. Rasbash, M. Yang, G. Woodhouse, H. Pan, D. Nu*ttall and S. Thomas, 1993. ‘A Multilevel* Analysis *o*f School Examination Results’, Oxford Review of Education, Vol. 19, No. 4, 1993.

Goldstein, H., and D. Spiegelhalter, 1996. ‘League Tables and their Limitations: Statistical Issues in Comparisons of Institutional Performance’, Journal of the Royal Statistical Society, Series A, Vol. 159, No 3, pp 385-443.

Goldstein, H. and S. Thomas, 1996. ‘Using Results as Indicators of School and College Performance’, Journal of the Ro*yal Statistical Society, Series A. Vol. 159, pp 149-164.*

Gray, J., D. Allnutt, J. Gardner, C. Blackham and B. Frost, 1995. ‘The Development of a National Framework for Estimating Value Added at GCSE A/AAS Level’, technical annex to GCSE *to GCE A/AS value Added - Br*iefing for Schools and Colleges, London: Department for Education and Employment.

Gray, J., H. Goldstein an*d D. Jesson, 1996. ‘Ch*anges and improvements in school’s effectiveness: trends over five years’, Research Papers in Education, Vol. 11, No.1, pp35-51.

Hutchinson, D., 1993. ‘School effectiveness studies using administrative data’, Educational Research, *Vol. 35, No. 1.*

Schools Curriculum and Assessment Authority, 1995. The Value Added National Project, Technical report: Secondary.

Thomas, S. and P. Mortimore, 1996. 'Comparison of VA models for secondary school effectiveness', Research Papers in Education, Vol. 11, No. 1, pp5-33.

Thomas, S., P. Sammons, P. Mortimore ad R. Smees, 1995a. ‘Differential Secondary School Effectiveness: Examining the size, extent and consistency of school and department effects on GCSE outcomes for different groups of students over three years’, Paper presented at the annual conference of the British Educational Research Association, Bath.

Thomas, S., P. Sammons, P. Mortimore, and R. Smees, 1995b. ‘Stability and Consistency in Secondary Schools’ Effects on Students’ GCSE outcomes over three years’ .School Effectiveness and School improvement (to appear).