Curriculum, Evaluation and Management Centre

PUBLICATION NO.22

SCHOOL EFFECTS AT A-LEVEL: GENESIS OF AN INFORMATION SYSTEM?

C. T. FITZ-GIBBON

1992

Chapter 5

School Effects at A Level: Genesis of an Information System?

Carol T. Fitz-Gibbon

INTRODUCTION

In England and Wales, externally set and marked examinations ('public examinations') are taken at the age of sixteen by a majority of pupils and at eighteen by the 20 per cent or so who choose to stay on for advanced work. At age sixteen, pupils now sit General Certificate of Secondary Education (GCSE) examinations, but at the time the project described in this chapter was undertaken, there was a two-tier system involving Certificate of Secondary Education (CSE) and General Certificate of Education (GCE) O (Ordinary) level examinations. The examinations taken at age eighteen were and still are GCE A (Advanced) levels. The last two years of school (equivalent to the US eleventh and twelfth grades) are called the 'sixth form'. The academic sixth form, those studying for A levels, although representing a small percentage of the school population, is the pool from which most future professionals are drawn and, as such, it is of undoubted importance in the field of formal education. Grades received at A level are used to screen applicants for admission to universities and polytechnics, thus making them significant factors in many careers.

The effects that schools have on A-level results are difficult to assess in any one year because of the relatively small numbers of students in each school, but given their importance as a major academic hurdle for entry to the professions they are worthy of close scrutiny and detailed monitoring. In schools, sixth forms are highly valued for the chance they give staff to teach to a high academic level and to teach highly motivated pupils. The academic content is substantial, as illustrated by the fact that a pass at A level in a subject may give as much as two years' credit in the subject in US universities. Moreover, the schools receive more funds for A-level students than for students lower down the school.

This chapter describes a research project on school effects at A level, i.e., in sixth forms. The project has since developed into a performance monitoring system (Fitz-Gibbon, 1989) supported by seven local education authorities (LEAs) and covering the north of England from the Tees to the Scottish border. In 1983 a letter was sent to schools in two LEAs inviting them to collaborate with a university department of education in evaluating A-level work. Since the letter was sent 'cold' (the researcher had no previous contacts with the schools) and in view of the fact that responses to questionnaires are often as low as 20 per cent, it came as a surprise when about 50 per cent of the schools responded positively, attended a meeting and joined the project. Moreover, all these schools have since continued to participate each year. Given this successful collaboration over a long period it may be worth identifying some operational features of the project before considering the findings.

Originally schools were, of course, being offered a free data collection and analysis service - a free information system. Nevertheless, the comparisons of one school against another might have been seen as potentially damaging or embarrassing. Anxieties on this account were probably reduced by two procedures. Firstly, the LEA did not receive the results. Permission to approach the schools was originally sought from the two LEAs involved, but the understanding was that the results would go in confidence directly to the schools. Secondly, each school chose a code-name so that although it received all the data and could compare its 'effectiveness' with that of other schools, the identity of the schools was not generally known. Each school knew only its own code-name. This need for confidentiality had to be considered in preparing the reports each year. For example, the size of the sixth form had to be omitted from tables because this datum by itself could identify a school.

Another feature which may have helped to ensure initial and continuing collaboration was the stringent efforts made to keep demands on school personnel to a minimum. Since 1984, the researcher or specially employed data collectors have administered all the tests and questionnaires. Schools need only make students available for a 70-minute session. This arrangement has the added benefit of ensuring high-quality data. The conditions for the test and the atmosphere in which the attitude-questionnaire is administered can be well standardized from school to school. In addition to having the measures administered by the same person in all schools, standardization is further improved by having the explanations and instructions pre-recorded on an audio-tape. Every candidate hears exactly the same explanation and instructions, and timing is consistent from school to school. These strenuous efforts to standardize the data collection are considered necessary because any differences introduced by administration procedures would be confounded with schools, the unit of analysis. Thus 'school effects' might appear which were in fact due to differences in the data collection procedures.

The project was named the COMBSE project, standing for Confidential, Measurement-Based, Self-Evaluation. The confidentiality has been explained above. The following explanations of 'measurement-based' and 'self-evaluation' were provided in 1983:

> **Measurement-based:** The aim of research is to *discover* relationships. We do not start with faith in any particular model of how schools should teach for A levels or what their policies should be. Instead we are concerned to find what *measurable* aspects of pupils, schools and teaching strategies *actually* relate to examination results and to pupils' satisfaction.
>
> **Self-evaluation:** We hope to find measurements that will help to predict A-level results and indicate *how much* difference is possibly attributable to various factors. However, it is quite certain that there will be many variations in A-level results which will not be explicable in terms of the data we have collected.

> Schools will be the final interpreters of their own particular set of results, knowing, as they do, far more than we can measure. However, by extending the areas that are measured, self-evaluation can be more informed.

It is just as well that, from the start, there was an emphasis on self-evaluation. Researchers could not acquire the detailed knowledge of the situation in each school that would be needed to interpret the data. The project represented, rather, 'monitoring' and the provision of information to those who could not only interpret it in the light of their full understanding of conditions in their school, but also take action on the trends suggested by the data, if that were appropriate. Providing clear and fair feedback to schools on their performance may be a feasible way to improve schools - letting schools improve themselves. Indeed, a re-analysis of the Hawthorne experiments suggested that feedback on performance was the source of 'the Hawthorne effect' (Parsons, 1974).

Table 5.1 indicates some of the data collected each year of the project. After the first year (some findings from which were reported by Fitz-Gibbon (1985)) the major sources of data have been questionnaires and ability tests administered to students in the 'upper sixth', one term before the A-level examinations.

In addition to investigation of the effects schools appeared to have on examination results, attitudinal variables were also reported each year. Three summated scales will be considered here: attitude-to-the-subject (mathematics or English), attitude-to-the-school and self-reported level of effort. The items contributing to these scales are listed in the Appendix to this chapter, in which the reliabilities of the scales are reported. In short, the study has provided schools with 'fair performance indicators' for examination results, attitudes to the school, attitudes to the academic subjects and levels of effort.

SOME FINDINGS

This chapter uses data from 1157 pupils taking A levels in the years 1983 to 1986. It is now of historical interest as documentation of the kind of research which led into a major performance monitoring system. Furthermore, while this was a small-scale study, as school effects studies go, it was nevertheless on the kind of scale at which a small, single LEA might collect data on A levels, so it was worthwhile asking what kind of stability and what kind of significant differences could be found on such a scale. The data were also used to generate hypotheses which could later be checked on larger samples.

Examination Effectiveness Scores

In order to evaluate examination results 'in context', i.e. controlling for intake characteristics, pupil-level prediction equations were needed, predicting A-level grades from information about each pupil. The predictor variables available were prior achievement at O level, ability tests and socioeconomic status. Since like predicts like, it was expected that prior achievement at O level would provide the highest correlation with achievement at A level and this was indeed the case.

Table 5.1 *Measurement topics in the COMBSE project*

Data collection procedure	*Major variables measured*	*Year 1* *1983*	*Year 2* *1984*	*Year 3* *1985*	*Year 4* *1986*	*Year 5* *1987*
Report from school	A level grades per pupil	✓	✓	✓		✓
	Examination Boards used		✓	✓	✓	✓
	Teacher experience with A level	✓				✓
	Class size	✓	✓	✓	✓	✓
	School size	✓	✓	✓	✓	✓
Report from LEA	School EPA classification			✓		
	School history	✓				
Test given to pupils	Pupil ability on APM[a]			✓		✓
	Pupil ability on AH6[b]		✓			✓
Pupil questionnaire	Pupil background (SES) + AVOC[c]	✓	✓	✓	✓	✓
	Pupil's attitude to school	✓	✓	✓	✓	✓
	Pupil's attitude to subject	✓	✓	✓	✓	✓
	Pupil's study habits	✓	✓	✓	✓	✓
	Classroom processes				✓	✓
Teacher interview	Classroom processes	✓				
	Very able pupils	✓		✓	✓	
	Experience of examination boards	✓				
Teacher questionnaire	Teacher-predicted A level grades	✓				✓
	Very able pupils			✓	✓	✓
	Classroom ethos			✓	✓	✓
	Textbooks used				✓	✓

[a]APM: Advanced Progressive Matrices (Raven, 1962).
[b]AH6: An NFER-published test for high ability pupils, developed by Alice Heim.
[c]AVOC: Average O level and CSE grade (the ILEA scale).

There are several different ways in which O levels might be used for the prediction of A levels, such as the number of passes, the number of 'good' passes or a sum-of-points scale, reflecting both the numbers of O-level examination successes and the levels of grade awarded. In this study the average grade obtained on whatever O levels were taken was found to be the best predictor of A-level grades. This could have been because the number of subjects for which a candidate entered was a matter on which different schools had different policies and these policies would thus represent a confounding factor if allowed to affect the intake measure. The average O-level grade represents several hours of academic testing on topics taught by several different teachers. It is a good index of general academic ability.

The correlations between each pupil's average O-level grade and A-level grades were 0.56 in English and 0.59 in mathematics. The average O-level grade obtained by an individual pupil will be referred to as the pupil's O-level GPA (grade point average) to utilize a concept familiar to American readers and to avoid confusions with group averages. It had been the intention to use multiple predictors but once this O-level GPA was entered into the prediction equation socioeconomic status measures did not make any additional statistically significant contribution to the prediction. The effects of home background have probably had their major impact before students appear in the sixth form. (However, Smith (1987) reported a similar finding for younger pupils, fifth-formers: social class showed little effect after ability had been considered.)

There were, however, two major problems with the use of O levels as predictors. One was that the effects of schools may already have been apparent in these very grades so that A-level effects will have been affected to an unknown extent by 'school effects' at O level. Schools which had been particularly effective with O levels might look poor at A level simply because the good O-level grades made it appear they were working with pupils with high abilities. Nevertheless, the effects calculated with O level grades as the predictor could justifiably be regarded as measures of 'improvement', 'change' or 'value added' between O level and A level. The other problem with the use of O-level grades in prediction equations for A levels was that in upcoming years O-level results would no longer be available as O levels were set to give way to other examinations. Since the project was to continue it was important to have some measures which provided comparability from year to year. Although not a problem for the present analysis this situation needed consideration for the future.

It was in order to deal with both these problems that ability measures were collected in 1984 (the AH6: Heim *et al.*, 1983) and in 1985 (the Advanced Progressive Matrices: Raven, 1965). The strongest correlations between these measures and A levels were considerably weaker than those provided by O levels, as would be expected since it is almost always the case that achievement measures predict achievement better than ability measures predict achievement. The AH6 verbal score had a correlation of 0.37 with A-level English and the Advanced Progressive Matrices had a correlation of 0.31 with A-level mathematics (see Table 5.2). Although weak, these correlations were as strong as or stronger than correlations generally found with socioeconomic status, a measure which is often accepted as a satisfactory covariate.

Table 5.2 *Correlations between ability measures and A-level grades*

Ability measure	*n*	*English grade*	*n*	*Maths grade*
APM[a]	122	0.21	122	0.31
AH6[b]	129	0.34	154	0.28
verbal		0.37		0.21
numerical		0.16		0.30
diagrammatic		0.22		0.17

[a]APM: Advanced Progressive Matrices (Raven, 1965).
[b]AH6: Alice Heim 6 (Heim *et al.*, 1983).

Having located O-level GPA as a moderately good predictor and ability tests as weak predictors of A levels, we could develop 'examination effectiveness scores' for each school for each subject for each year, using either predictor. The regression lines used for reports to the schools each year were based on the data for that year. For this chapter, results have been pooled across the four years 1983 to 1986 and a single regression line used. (The small sample sizes for each of the ten schools precluded consideration of separate regression lines per school per year). For each school, the average of the pupils' residuals was designated the examination effectiveness score for the school.

Did the two predictors, prior achievement and ability, yield similar results? Yes, to a large extent. The correlations between the two types of residuals for individual pupils were 0.80 for both the AH6 and the APM with English candidates, and 0.85 and 0.75 respectively for mathematics candidates.

Levels of Analysis and Levels of Aggregation

As Plewis commented in a symposium on 'Statistical modelling issues in school effectiveness studies' (Aitken and Longford, 1986), 'it should not be forgotten that pupils are taught by teachers not by schools'. Many would argue that the major impact on the pupil is the particular teacher, so that it is at the teacher level that one should expect to find the maximum 'effects' of schooling.

In the COMBSE study the decision was made from the start to examine effects subject by subject, with English and mathematics chosen initially as subjects which represented the science/arts divide and which had relatively large enrolments. As shown in Table 5.3, this decision seems justified by the data. In an analysis of variance with mean O-level grade as a covariate, the effect of subject (English or mathematics) was highly significant *and* the interaction of subject with school was highly significant, indicating that different schools obtained good results in different subjects. This finding runs contrary to that of Willms and Cuttance (1985), in their study of Scottish leaving qualifications among fifteen schools. Willms and Cuttance found schools that were effective in English tended also to be effective in mathematics. Perhaps the younger age group accounts for the different finding. The implication of Table 5.3 is that, for A-level effects, to aggregate results across subjects would be misleading and would result in important variations being overlooked.

In short, it appeared to be the case that schools were not uniformly effective

Table 5.3 *The effects of subject (English or maths) and school attended on A-level grades achieved, with average O-level grade controlled*

Source of variation	*Sum of squares*	*DF*	*Mean square*	*F*	*Significance of F*
Average O-level grade	981.764	1	981.764	443.341	0.000
Subject	82.367	1	82.367	37.195	0.000
School	36.476	9	4.053	1.830	0.059
Subject school	57.717	9	6.413	2.896	0.002
Residual	2079.384	939	2.214		
Total	3232.474	959	3.371		

in getting A-level grades but might be effective in English and not effective in mathematics, or vice versa. If this kind of variation in effectiveness from department to department within a school is a general phenomenon then it casts doubt on the wisdom of parents or researchers trying to locate the best *school*. It might also be taken as an indication that efforts to improve education must be made *within schools*, department by department, rather than by setting schools in competition.

Since data in the COMBSE project were collected subject by subject, was the effectiveness of individual A-level teachers being investigated? No. Almost all the A-level classes were taught by a team of teachers so that the effects of a single teacher on A-level results could not be examined. The unit of aggregation in the COMBSE study is most properly thought of as the school *department*. Although individual teaching groups within schools (i.e. classes) were identifiable each year, changes in the composition of the teaching team from year to year meant that no information could be obtained on the stability of effects among classes. However, given the yearly feedback of effectiveness data, the schools themselves could make interpretations based on their intimate knowledge of conditions and personnel in the department each year. This was the 'self-evaluation' component of the COMBSE project. It bears repeating that an outsider collecting and analysing data will not have the detailed, qualitative knowledge of conditions that is available to personnel working in the schools. Furthermore, staff appraisal requires more complex information than that provided by objective data analyses.

The Magnitude of Differences in Effectiveness

The 'examination effectiveness score' employed in this study may be defensible but it fails to convey a sense of whether or not the differences between schools were sufficiently large to be of concern. Did it matter much whether a pupil attended a school with low or high examination effectiveness? Table 5.4 shows the actual distribution of the grades which were obtained by schools with the lowest and schools with the highest effectiveness scores.

For both English and mathematics the probability of getting an 'A' or a 'B' was approximately twice as high in the most effective schools as in the least effective. In the most effective schools the chance of a 'Fail' or the derisory

Table 5.4 *Percentage of pupils who obtained various grades at the least and most effective schools*

Schools	*n*	*A-level grade*						
		Fail	*O*	*E*	*D*	*C*	*B*	*A*
English								
Three least effective	75	33	20	24	5	8	7	3
Two most effective	90	3	24	21	13	12	15	11
Mathematics								
Two least effective	154	8	31	15	20	12	9	5
Three most effective	157	10	17	18	17	13	12	13

n: Number of pupils.

O-level pass was half that in the least effective schools in English and about 70 per cent as great in mathematics.

Causal Attributions for 'Effectiveness'

Although the teaching-team approach to A levels provided a comfortable reason to stay clear of issues of teacher effectiveness, it left open the question of how effectiveness might be explained or understood. What causes a school department to be effective? One seeks more than simply the statement that there are effects of various magnitudes. Explanations of why there are effects are needed.

There are many ways forward on this question, most of which will require, initially, the collection of qualitative information about teaching practices and the management of the school department. This information should be collected on site *by someone who is unaware of the effectiveness data.* Meanwhile, from a distance, a start can be made on collecting some evidence of associations between classroom processes and effectiveness scores. Since 1986 the project has collected process data about what happens in classrooms, using questions some of which matched those in Gray *et al.* (1983). By regarding pupils as 'raters' and using the class as the unit of analysis, some scales which have reasonably high degrees of inter-rater consistency and which differentiate between classrooms have been developed. A scale assessing the amount of pupil talk was constructed by asking about the frequency of such activities as 'discussion in groups', 'working in pairs', 'presenting your work to the class' and 'listening to another student present work to the class'. A visit to several schools by someone unaware of the data on the pupil-talk scale produced a glowing report of the amount of pupil participation in lessons in exactly the English department which had the highest score on the pupil-talk scale, thus providing some reassurance of validity for the scale as a measure of classroom events. This process variable did not, however, correlate significantly with examination effectiveness scores. Given the small sizes of samples for each year it is important to collect several years' data before drawing conclusions. Work of this kind, relating statistical descriptors and ethnographic descriptions, is urgently needed.

Another scale, 'examination emphasis', was moderately reliable for English but for mathematics there was more variation between pupil reports within classrooms than between classroom averages, making an unsatisfactory scale. Again the importance of dealing with different academic subjects differently is emphasized.

This search for processes that relate to effectiveness scores is important for two major reasons. One is that if effective processes can be located, it may be possible for these to be adopted elsewhere with good results. The search for effective processes is the search for 'alterable variables' (Bloom, 1979) or 'tractable variables' (Willms and Cuttance, 1985). In contrast, locating effective teachers would still leave open the question of how they were effective and whether or not they could be copied. A question on which there is still very little evidence is the amount of variance in effectiveness accounted for by non-imitatable teacher variables, such as personality or charisma, and the amount which is accounted for by processes which could be adopted by most teachers. Only experiments will even begin to resolve this issue.

Another reason for the importance of the search for process variables which have some predictive validity is that they might be our only chance of explaining the unstable outcomes from year to year, from school to school - for the outcomes *were* unstable, as will be discussed below. Faced with instability in the data, one has to ask if effectiveness was 'really' varying from year to year or whether the instability was almost entirely due to poor models and/or variables with low reliability and validity. Finding process variables which 'explain' some of the variations in effectiveness is thus an important test of the models used in school effectiveness research.

Stability of Effectiveness Scores from Year to Year

Table 5.5 indicates that there were no trends in examination effectiveness from year to year (no significant main effects for year) but there were highly significant interactions between year and school, indicating different schools were effective from year to year. (Again this emphasizes what seems to be the futility of a search for effective *schools*: the results would depend upon which year's data were examined.)

One way to index the stability from year to year for several variables, including examination effectiveness, was to compute KR-20 as a measure of the reliability of the school means from year to year (McKennell, 1970). Results of this analysis are shown in Table 5.6. This index of reliability is perhaps better called an index of relative stability. If the schools in the sample maintained roughly the same rank order from year to year and if the variation within schools was small compared to the variation between schools, then KR-20 would be large.

There was considerable relative stability among schools in the ability of pupils staying on into the sixth form (0.75 for English and 0.87 for mathematics). In other words, their intakes were stable, a finding consistent with the stable population patterns associated with north-east England. In the sixth form, however, many aspects changed from year to year so that instability might be expected and was indeed evident, with low relative stability on mean raw A-level scores and on examination effectiveness scores (mean residual gains) with the

Table 5.5 *ANOVA: the effects of cohort (data year) and school on A-level grades adjusted for O-level GPA*

Source of variation	*Sum of squares*	*DF*	*Mean square*	*F*	*Significance of F*
English					
Covariate (O-level GPA)	450.947	1	450.947	218.829	0.000
Data year (1983, '84, '85 or '86)	3.742	3	1.247	0.605	0.612
School	30.130	9	3.348	1.625	0.106
Data year by school	114.533	24	4.772	2.316	0.001
Residual	799.561	388	2.061		
Total	1398.913	425	3.292		
Mathematics					
Covariate (O-level GPA)	618.995	1	618.995	295.026	<0.001
Data year (1983, '84, '85 or '86)	2.799	3	0.933	0.445	0.721
School	76.230	9	8.470	4.037	0.000
Data year by school	96.397	25	3.856	1.838	0.008
Residual	1038.561	495	2.098		
Total	1832.981	533	3.439		

Table 5.6 *KR-20 for school mean values from years 1984, 1985 and 1986*

Variable	*English*	*Maths*
O-level GPA	0.75	0.87
Raw A-grade	0.26	0.35
Exam. effectiveness score	0.20	0.61
Attitude-to-the-subject[a]	b	0.47
Attitude-to-the-school[a]	0.19	0.20
Reported effort[a]	0.50	b

[a]Summated scales described in the Appendix.
[b]Essentially zero. More variation within than between.

exception of mathematics effectiveness. Among the changes which might explain this variation in the data were changes in teachers and, in a few cases, in examination boards or syllabuses.

The Pulling Power of Some Mathematics Departments

Another variable which came close to the O-level GPA in being stable from year to year was the ratio of the number of candidates choosing mathematics to the number of those choosing English. Nationally this ratio was about 1.4; that is, about 40 per cent more candidates entered for mathematics than entered for English. The schools in the COMBSE study showed a considerable range, from 0.6 to 4.5 in 1986, for example. Some mathematics departments were attracting many more entrants than English departments. This was dubbed the 'pulling power' of the mathematics department. The pulling power scores showed a KR-20 value of 0.82, indicating that mathematics departments were fairly consistently popular or unpopular.

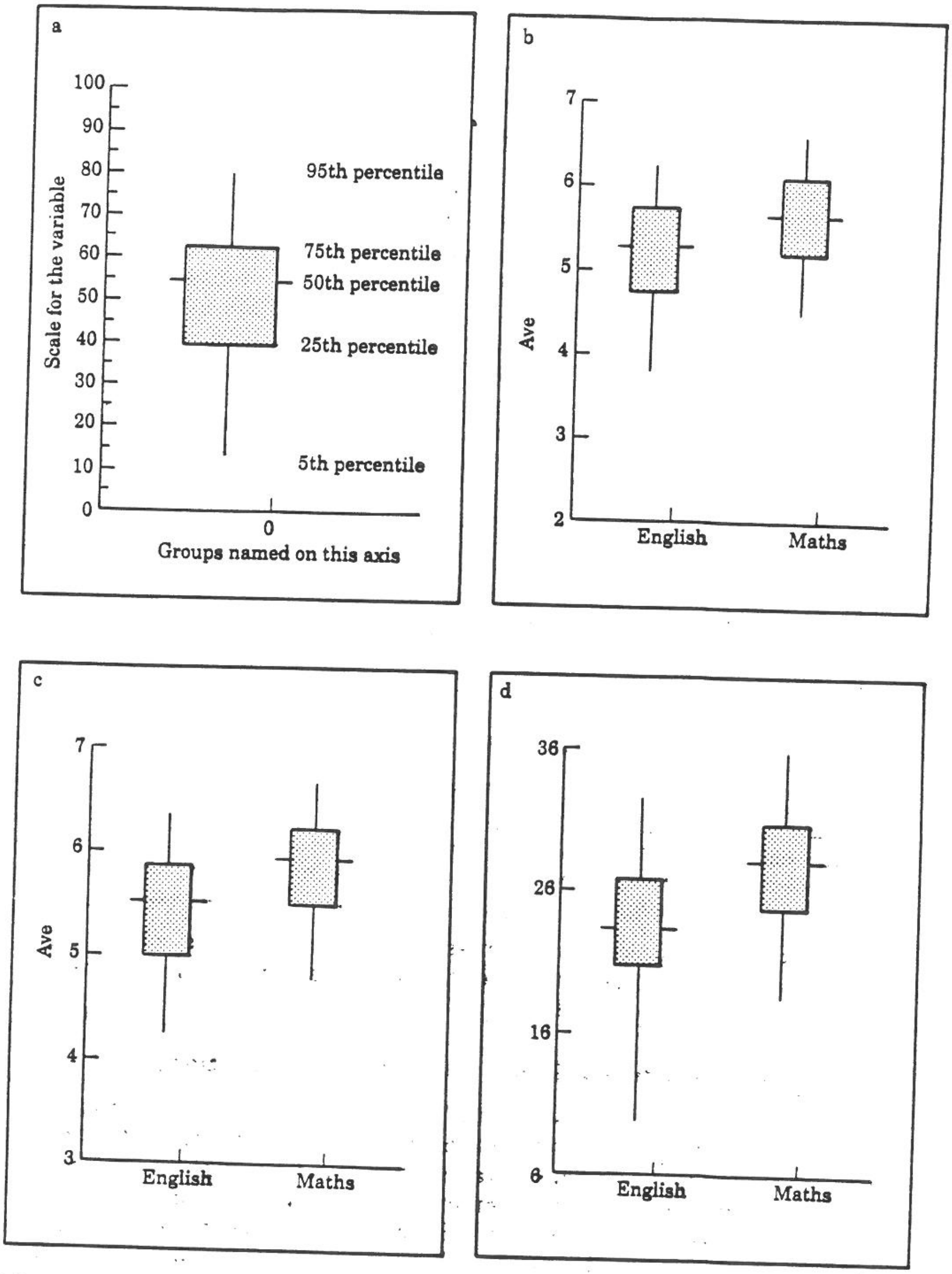
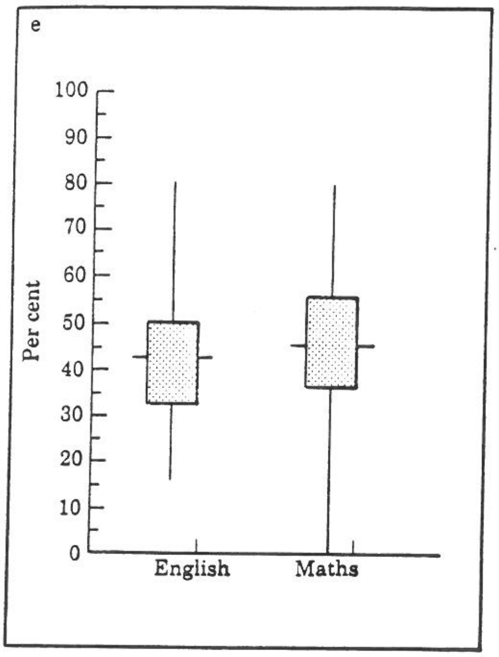
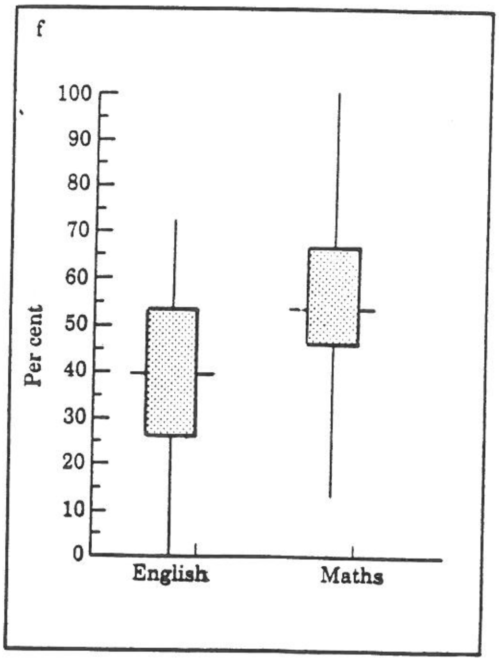
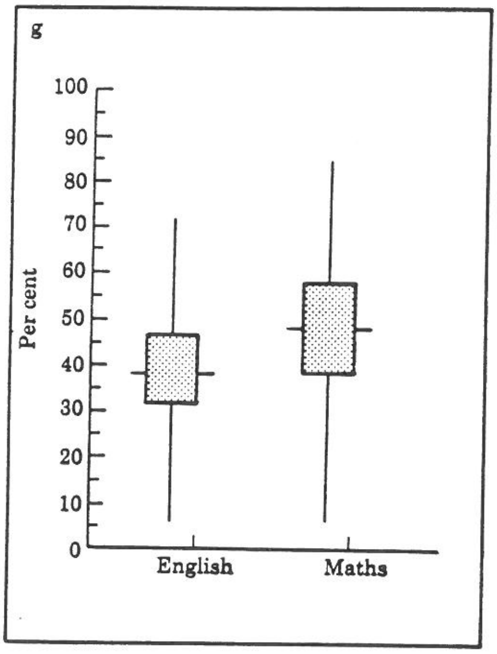


Figure 5.1 *Attainments of candidates for A-level English and mathematics on prior achievement and ability measures. a, Key. b, Mean O-level grades of all candidates. c, Mean O-level grades of successful candidates. d, Advanced matrices. e, AH6 verbal. f, AH6 numerical. g, AH6 diagrammatic*


e
100
90
80
70
60
50
40
30
20
10
0
Per cent
English
Maths


f
100
90
80
70
60
50
40
30
20
10
0
Per cent
English
Maths


g
100
90
80
70
60
50
40
30
20
10
0
Per cent
English
Maths

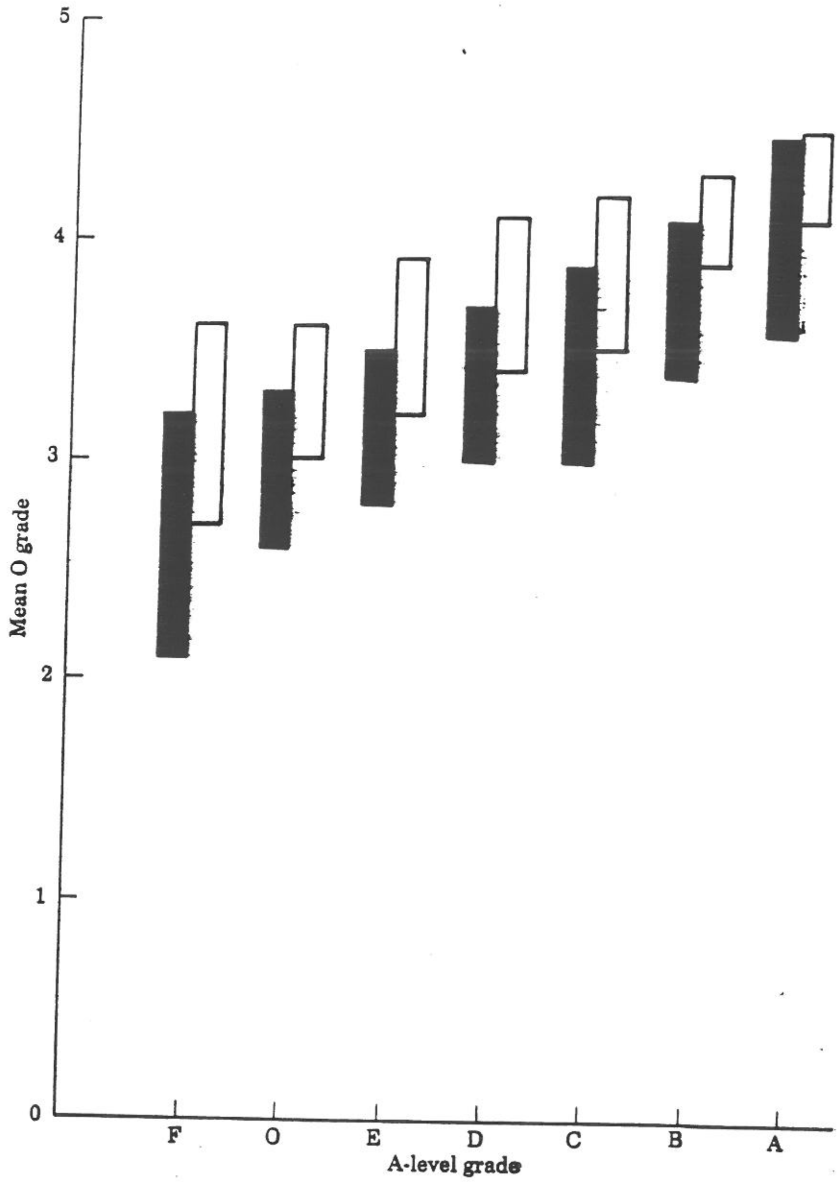


Figure 5.2 *O-level GPA interquartile ranges for groups that attained various A-level grades in English (filled bars) and mathematics (open bars)*

The Higher Ability of Mathematics Candidates

It was clear that mathematics generally attracted more able pupils. The ability tests reported above served to confirm a pattern which was clear from examination of O-level GPAs: on average mathematics candidates were substantially more able than English candidates. Graphs for O-level GPAs, the Advanced Progressive Matrices and the AH6 subscales are presented in Figure 5.1.

Other studies have arrived at the same conclusion (Smithers *et al.*, 1984; Smithers and Robinson, 1987). Whether one looks at O-level GPA, ability as measured by the various subscales of the AH6 or the scores (of another cohort) on the APM, the result replicates again and again: mathematics candidates were substantially more academically able. On mean O-level grades the lower quartile score for successful mathematics candidates was about the same as the median score for successful English candidates. The 'effect size' or standardized mean difference was 0.67, implying a difference of about two-thirds of a standard deviation. On the APM the effect size was even larger: 1.02. On the AH6 the effect size was 0.62 on the full scale.

Was a 'D' in mathematics worth a 'B' in English? Mathematics candidates were more able than English candidates but a similar proportion failed. This implied that the mathematics examination at A level was more difficult than the English examination at A level. Under such an examining system, making efforts to attract the hesitant, borderline pupil into mathematics A level might have been dangerous. Schools had to warn their students that they were more likely to fail mathematics than English, other things being equal. The bars in Figure 5.2 represent the range from the lower to the upper quartile on the distribution of O-level GPA for pupils who attained the indicated grade at A level. In other words, each bar represents the spread of average O-level grades for the middle half of the group of pupils. It can be seen that the range of abilities of the middle 50 per cent of candidates who obtained a 'D' in A-level mathematics was almost exactly the same as the range for those who obtained a 'B' in A-level English.

This difference in examination difficulty at A level is not a law of nature; it is a policy decision and one which needs reconsideration. Should mathematics be left as so much more difficult than English? Do we need to push people out of mathematics so early? Should we then be surprised that many otherwise well educated people feel mathematically illiterate, that there is a shortage of people with mathematical competencies in so many professions and, in particular, a shortage of mathematics teachers?

SENSITIVITY TO SCHOOL EFFECTS

Table 5.6 presented indices of stability of means from year to year and it was notable that there was considerably more stability in effectiveness in mathematics than in English (0.61 as opposed to 0.20). A slightly different question to ask of the data relates to the proportion of variance explained by the variable 'school' (or school department in the present study). This 'proportion-of-variance-accounted-for' measure can be thought of as indicating the effects which schools can have relative to the large amount of variation in the data owing to differences among pupils.

Discussions of the proportion-of-variance-accounted-for by schools have generally set the figure quite low. For a sample of comprehensive schools Aitkin and Longford (1986, p. 15) reported a figure of 7 per cent for effects on O-level achievements as indexed by a sum-of-points scale. Willms (1987), employing the Scottish School Leavers Surveys of 1977, 1981 and 1985, reported a figure of about 10 per cent.

In addition to examining the proportions of variance in examination effectiveness, one could also ask to what extent schools had effects on the three scales mentioned earlier: attitude-to-the-subject, attitude-to-the-school and effort. Except for attitude-to-mathematics, these attitude scales were not significantly correlated with prior achievement or ability measures. Consequently, if there were to be any control for differences in school intakes it would have to be by means of socioeconomic status (SES). The somewhat surprising finding here was that more positive attitudes were associated not with higher SES but with lower SES of pupils. *Post hoc* explanations spring to mind. It could be that higher SES students in the sixth form were more inclined to be critical and students from lower SES backgrounds were more inclined to be appreciative. Again these hypotheses need exploration by qualitative methods. For our purposes here, SES was used as a covariate to control for some intake differences when examining attitudes.

Susceptible Outcomes

There may be some outcomes that can be little affected by anything schools can do, whereas schools may have a large impact on other outcomes. Which outcomes are, or are not, sensitive to school effects? *A priori*, one might hypothesize that mathematics would be more sensitive to schooling than English. Poor instruction in mathematics may have worse effects than poor instruction in English and likewise particularly good instruction in mathematics might make more of an impact than particularly good instruction in English, particularly on achievement. Another way of putting this hypothesis is that pupils are more dependent on schooling for learning mathematics than for learning English literature.

The proportion-of-variance-accounted-for was computed from one-way ANOVAs treating schools as a random factor. The results are presented in Table 5.7. In Figure 5.3 this proportion-of-variance-accounted-for by schools has been graphed for each of the four outcomes, with separate bars for English and mathematics.

In the present data the proportion-of-variance-accounted-for by schools was about 6 per cent in mathematics as compared to only 1 per cent in English. It appeared that among mathematics candidates achievement and attitude-to-the-subject were far more susceptible to school effects than they were among English candidates, as hypothesized above. One could say that in this sample the proportion of variance accounted for was about six times larger for mathematics than for English. When we examine the outcome most closely related to the school, however, namely the pupil's attitude to the school, the proportion of variance was very large in English (over 15 per cent) although very small in

Table 5.7 *Susceptibility to school effects among four dependent variables*

DF (N – 1)	*English or Maths*	*Dependent variable*	*Controlling for*	*Proportion of variance accounted for (random model)*	*F*	*p*
A-level grades with no control variables						
425	English	A-grade		0.384	6.11	<0.001
533	Maths	A-grade		0.102	2.54	0.01
Four dependent variables as plotted in Figure 5.2						
425	English	A-grade	GPA	0.010	1.41	0.18
533	Maths	A-grade	GPA	0.057	4.04	<0.001
207	English	Att. to sub.	SES	0.004	1.08	0.38
251	Maths	Att. to sub.	SES	0.111	3.99	<0.001
289	English	Att. to sch.	SES	0.155	5.83	<0.001
346	Maths	Att. to sch.	SES	0.092	4.28	<0.001
280	English	Effort	SES	0.016	1.43	0.17
341	Maths	Effort	SES	0.000	0.90	0.52

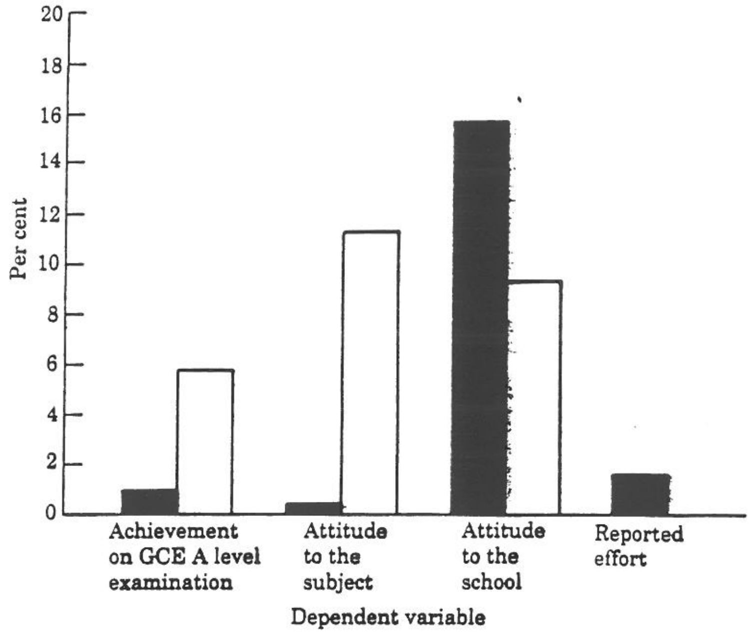


Figure 5.3 *School effects: percentage of variance on the random effects model for English (filled bars) and mathematics (open bars)*

mathematics. Are English candidates more aware of, or more responsive to, school 'atmosphere' than mathematics candidates?

Turning to the question of school effects on behaviours, examination of the self-reports of effort showed exceedingly small effects. Yet teachers try hard to influence the effort that pupils expend. These findings suggest another hypothesis: if effort is little affected but achievement (in mathematics) is, perhaps good results depend upon the quality of the work done, not the quantity of time expended on the work. The more effective teachers do not set more work than the less effective teachers, but set better work or teach better. But this is only a hypothesis, going beyond the data available.

Carroll, in his seminal 'Model of school learning' article, commented that we often think that motivation (effort or perseverance) is a behaviour that can be influenced while academic achievement is determined by fixed abilities. The facts might be otherwise.

> 'Aptitude' is regarded as relatively resistant to change, whereas it is the hope of the psychologist that he can readily intervene to modify 'perseverance'.... To some extent, this feeling is justified not only by logic but also by research findings - by the research on the apparent constancy of the IQ.... On the other hand, if aptitude is largely a matter of prior learnings, it may be more

> modifiable than we think. Whereas, conversely, some kinds of clinical findings suggest that motivational characteristics of the individual may be much harder to change than one might think.
> (Carroll, 1963, p. 731)

From the current data this would seem to be the case for mathematics: the variance attributable to the effects of schools was approximately 0 per cent for effort as opposed to 5.7 per cent for achievement. In English both effects were small (1.6 per cent for effort and 1 per cent for achievement).

Susceptible Sub-groups of Pupils

In addition to examining the extent to which various outcomes are susceptible to schooling, one might ask whether schools have different amounts of impact on different groups of pupils (cf. Reynolds and Reid, 1985).

Do school effects differ with respect to the ability level or socioeconomic status of pupils? We might hypothesize that pupils who are either more able or from more affluent backgrounds are in a better position to compensate for a poor instructional programme - they have more resources either within themselves or at home on which they can draw if they are not learning at school. Indeed there were several schools in the COMBSE sample in which pupils reported the use of private tutors. Given these personal or financial resources, were high ability or high SES pupils less susceptible to school effects than less able or lower SES pupils? Were low ability and low SES pupils more susceptible to school effects because they were more dependent on schools, having fewer alternative resources?

Table 5.8 presents the proportions of variance accounted for when sub-groups of high and low SES, and high and low ability, were examined. The hypothesis that pupils of low ability or low SES would be the more 'vulnerable' did *not* appear to be supported by the data. At A level, the schools appeared to have had more effect on more able pupils. This conclusion applied to achievement, attitude-to-the-school and attitude-to-the-subject but not to effort.

Post-hoc, what are the possible explanations? One consideration is the nature of the A-level scale. A levels show little differentiation at the lower end of the ability continuum: nationally 50 per cent received an E or failed. There was more differentiation at the upper range, in the D, C, B and A grades. It is in this range that the effects were detectable. One might also suggest that schools have the greatest effects on pupils for whom the curriculum is best suited. The A-level curriculum was best matched to the abilities of the more able pupils. However, while not supported by the present dataset, the 'disadvantaged are more dependent on schools' hypothesis does receive support from data presented by Cuttance in this volume. This serves as a warning not to assume that effects at one level of the educational system have the same patterns as effects at another level.

While the findings on the sub-groups of pupils will need more replication, being based on small numbers and yielding some surprising results, the illustration that mathematics is more sensitive to schooling than English is so much in

Table 5.8 *Susceptibility to school effects among sub-groups*

DF *(N – 1)*	*Sub-group*	*Dependent variable*[a]	*Proportion of variance accounted for (random model)*	*F*	*p*
English sample broken down by SES and ability					
130	Low SES	A-grade	–0.064	0.26	0.98
159	High SES	A-grade	0.027	1.39	0.20
279	Low ability	A-grade	0.009	1.24	0.26
145	High ability	A-grade	0.015	1.17	0.31
68	Low SES	Att. to sub.	0.016	1.11	0.37
138	High SES	Att. to sub.	0.019	1.26	0.27
128	Low ability	Att. to sub.	–0.010	0.87	0.55
78	High ability	Att. to sub.	0.135	2.05	0.05
128	Low SES	Att. to sch.	0.024	6.00	<0.001
160	High SES	Att. to sch.	0.310	7.34	<0.001
182	Low ability	Att. to sch.	0.167	4.48	<0.001
106	High ability	Att. to sch.	0.178	2.84	0.005
124	Low SES	Effort	0.079	2.00	0.04
155	High SES	Effort	–0.028	0.63	0.77
270	Low ability	Effort	0.039	2.07	0.03
143	High ability	Effort	–0.022	0.76	0.65

Maths sample broken down by SES and ability					
155	Low SES	A-grade	− 0.025	0.63	0.77
187	High SES	A-grade	0.061	2.26	0.02
231	Low ability	A-grade	0.049	2.14	0.03
301	High ability	A-grade	0.067	2.98	0.002
81	Low SES	Att. to sub.	0.013	1.10	0.37
169	High SES	Att. to sub.	0.154	3.77	0.002
110	Low ability	Att. to sub.	0.171	3.17	0.002
140	High ability	Att. to sub.	0.138	3.03	0.002
154	Low SES	Att. to sch.	0.023	1.34	0.22
191	High SES	Att. to sch.	0.174	4.27	< 0.001
144	Low ability	Att. to sch.	0.054	1.78	0.08
201	High ability	Att. to sch.	0.129	3.69	< 0.001
150	Low SES	Effort	0.023	1.34	0.22
190	High SES	Effort	0.019	1.33	0.22
229	Low ability	Effort	− 0.012	0.73	0.68
304	High ability	Effort	− 0.002	0.95	0.48

Negative variance proportions imply that the model did not hold. They are usually set to zero.
[a]The same variables were controlled for as in Table 5.7.

conformity with expectations and intuitions that it seems likely to become a robust finding. It has implications that may not be uniformly welcomed: efforts to improve achievement may be more likely to result in measurable improvements if they are concentrated on mathematics rather than on English teaching, at least in the sixth form. On the other hand, since there is national concern over levels of mathematical competencies, findings that schools do affect mathematics achievement must be welcome.

CONCLUSIONS

Two kinds of conclusion arise from research: those which can be fairly said to have arisen directly from the data and those at which the researcher has arrived as a result of conducting the research project. The latter, based it might be said on 'experiential' rather than formal learning, must be regarded as closer to opinions than to findings, but they are often important.

Conclusions Arising from the Data

From the data collected over four years from ten comprehensive schools, it would seem to be the case that:

1 Schools varied significantly and substantially in the level of A-level grades their students attained, even after taking account of prior achievement or ability.

2 Schools that appeared to be effective in getting good grades in English were not necessarily the same ones that were effective at getting good grades in mathematics.

3 There was considerable variation from year to year in the effectiveness of English and mathematics A-level work. Since it is known that there were changes in the schools in such areas as teachers, examination boards and syllabuses, this year-by-year fluctuation would be expected. More data would be needed to explain some of the year-by-year fluctuation, including qualitative data and classroom process data.

4 Socioeconomic status was very weakly correlated with achievement in A-level work. An ability measure or a prior achievement measure is therefore a necessity for use as a co-variate in data at this level of the educational system.

5 In this sixth-form sample, socioeconomic status correlated negatively with attitude-to-the-school: pupils from homes with parents in professional jobs were less satisfied than students whose parents were in lower-status jobs. This finding is contrary to the pattern generally expected in younger pupils.

6 Mathematics appeared to be more sensitive to school effects than English. The variance in examination results which could be attributed to schools was about six times larger among mathematics students than among English students.

7 School effects on expressed attitudes were stronger than their effects on achievement, in terms of variance accounted for.

8 School effects on the level of effort reported by pupils were surprisingly small and not statistically significant. This finding contrasts with the feeling teachers have that they affect the effort students make.

9 Candidates for A-level English were considerably less able, as a group, than candidates for A-level mathematics. A D in mathematics A-level tended to be awarded to candidates with the same range of ability as those obtaining a B in English.

Conclusions Arising from the Experience of the Research

The following conclusions cannot be said to arise directly from the data but rather from reflection upon the data and upon the experience of collecting and analysing the data and feeding them back to schools.

1 The school department is the desirable unit on which to focus as a first level of aggregation in school effects studies. Because there are significant differences between the effectivenesses of departments within the same school, aggregation to the level of school will mask important effects. Furthermore, the school department is a unit which is *managed*. Information on effectiveness can actually be used by a school department. It is less clear how to go about improving a whole school.

2 School personnel need assistance in interpreting and using school effectiveness data. It is not enough simply to send out the reports each year.

3 As the quantity of data becomes larger, more sophisticated analysis strategies will become possible and desirable, utilizing computer programs written especially to handle hierarchical, nested data (Aitkin and Longford, 1986; Goldstein, 1985; Raudenbush and Bryk, 1986).

Finally, the most important conclusion drawn was:

4 The logical outcome of school effectiveness research is the creation of monitoring systems to supply schools with regular measures of effectiveness. The monitoring systems must provide fair indicators of performance, not only on cognitive goals but also on other outcomes of concern. This information must be collected in collaboration with school departments and reported back to school departments. They are close enough to the events to interpret the data and take action where necessary. If this monitoring is to be fair it must develop from the base of information and experience which is being slowly built up by research into school effectiveness.

APPENDIX: THE SCALES USED

The A-level Scale

This was a modification of the scale used by UCCA. It is extended at the lower end to distinguish between an O-level pass (coded 0) and an outright Fail (coded −1). (These grades became roughly 'N' and 'U' in 1987).

5 points for an A
4 points for a B
3 points for a C
2 points for a D
1 point for an E
0 points for an O
− 1 point for an F

Attitude to the Subject Scale

Six items were summed to yield the attitude-to-English scale or the attitude-to-mathematics scale. These were all five-point Likert-type items on the questionnaire. Stated positively, these items were:

17 Not finding it hard to get down to work in the subject.
18 Looking forward to lessons in the subject.
19 Liking examinations in the subject.
20 Thinking about the subject outside class.
21 Not regretting having chosen the subject.
22 Preferring the subject to others being studied.

In the 1986 data the reliabilities of this attitude-to-the-subject scale were:

0.72 in English	(n = 175)
0.76 in mathematics	(n = 245)

Attitude-to-the-school Scale

Six items were summed to yield the attitude-to-the-school scale, all from Likert-type items on the questionnaire. Stated positively, these items were:

1 Liking school.
2 Liking lessons.
3 Liking the teachers.
4 Feeling one was being treated like an adult.
5 Thinking the atmosphere in the school was good for sixth formers.
6 Reporting one would recommend the school to others.

Previous work has shown that sixth formers place considerable importance on being treated in an adult fashion. The reliabilities of this scale were:

0.83 in English $(n = 182)$
0.77 in mathematics $(n = 247)$

Effort Scales

'Effort' was assessed by summing the responses to the following items on the questionnaire:

1 Time spent per week on homework, categorized on a six-point scale.
2 Time spent per evening on homework for all subjects.
3 Not doing homework while watching TV.
4 Getting work in on time.
5 Doing more than just the required amount of work.
6 Working hard.
7 Being one of the hardest workers in the class.

A person obtained a high score on the effort scale, therefore, to the extent that he or she reported spending time on the subject and on homework in general, not watching TV while doing homework, getting work in on time, doing more than only what was required, working hard and being one of the hardest workers in the class. The reliabilities of the two scales were as shown below:

0.65 in English $(n = 166)$
0.65 in mathematics $(n = 225)$

Socioeconomic Status Scale

Socioeconomic status was assessed in the following way:

I Professional: accountant, doctor, lawyer, clergyman, etc. coded 6
II Intermediate: Member of Parliament, nurse, manager, etc. coded 5
III Skilled non-manual: clerical worker, sales representative, etc. coded 4
III Skilled manual: bus driver, butcher, bricklayer, etc. coded 3
IV Partly skilled: barman, fisherman, postman, etc. coded 2
V Unskilled: kitchen hand, labourer, office cleaner, etc. coded 1

The reliability of the SES index was:

0.55 in English $(n = 122)$
0.57 in mathematics $(n = 148)$

ACKNOWLEDGEMENTS

The author wishes to thank the Department of Education and Science for a small grant, Peter Clarke for collecting the 1987 data so effectively, Patrick Eavis for 'blind' visits to English departments and most useful descriptions thereof, Tony Edwards and Peter Cuttance for helpful comments on a draft and, especially, all the participating schools and colleges.

REFERENCES

Aitken, M. and Longford, N. (1986) 'Statistical modelling issues in school effectiveness studies', *Journal of the Royal Statistical Society, Series A*, 149 (1), 1-43.

Bloom, B.S. (1979) *Alterable Variables: The New Direction in Educational Research*. Edinburgh: Scottish Council for Research in Education.

Carroll, J.B. (1963) 'A model of school learning', *Teachers College Record*, 64, 723-33.

Fitz-Gibbon, C.T. (1985) 'A-level results in comprehensive schools: the COMBSE project, year 1', *Oxford Review of Education*, 11 (1), 43-58.

Fitz-Gibbon, C.T. (1989) 'Using performance indicators: educational considerations', in Levacic, R. (ed.) *Financial Management in Education*. Milton Keynes: Open University Press.

Goldstein, H. (1985) 'Multi-level mixed linear model analysis using iterative generalised least squares', *Biometrika*, 73 (1), 43-56.

Gray, J., McPherson, A.F. and Raffe, D. (1983) *Reconstructions of Secondary Education: Theory, Myth and Practice since the War*. London: Routledge & Kegan Paul.

Heim, A.H., Watts, K.P. and Simmonds, V. (1983) *Manual for the AH6 Group Tests of High-level Intelligence*. Windsor: NFER.

McKennell, A.C. (1970) 'Attitude Scale construction', in O'Muircheartaigh, D. and Payne, G. (eds) *The Analysis of Survey Data*. Vol. 1: *Exploring Data Structures*. Chichester: John Wiley.

Parsons, H.M. (1974) 'What happened at Hawthorne?' *Science*, 183, 922-32.

Raudenbush, S. and Bryk, A.S. (1986) 'A hierarchical model for studying school effects', *Sociology of Education*, 59, 1-17.

Raven, J.C. (1965) *Advanced Progressive Matrices. Sets I and II*. London: H.K. Lewis.

Reynolds, D. and Reid, K. (1985) 'The second stage: towards a reconceptualization of theory and methodology in school effectiveness research', in Reynolds, D. (ed.) *Studying School Effectiveness*. Lewes: Falmer Press.

Smith, I. (1987) 'Educational and vocational opportunities at 16+', *ESRC Newsletter*, 61, 19-20.

Smithers, A.G. and Robinson, P. (1987) 'Mixing A levels: hard choice or soft option?', Paper presented to the Thirteenth Annual Conference of the British Educational Research Association, Manchester, 2-5 September.

Smithers, A.G., Collings, J.A. and McCreesh, F.J. (1984) 'The growth of mixed A level courses', *Research in Education*, 32, 1-16.

Willms, J.D. (1987) 'Differences between Scottish educational authorities in their examination attainment', *Oxford Review of Education*, 13 (2), 211-32.

Willms, J.D. and Cuttance, P. (1985) 'School effects in Scottish secondary schools', *British Journal of Sociology of Education*, 6 (3), 289-306.