



THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2000/52

The Comparative Evaluation of GCSE Value-Added Performance
by Type of School and LEA

by

David Jesson

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD



CENTRE FOR PERFORMANCE
EVALUATION AND
RESOURCE MANAGEMENT

**THE COMPARATIVE EVALUATION OF
GCSE VALUE-ADDED PERFORMANCE
BY TYPE OF SCHOOL AND LEA**

by

DAVID JESSON*

*The author is Associate Director, and Honorary Visiting Professor, in the Centre for Performance Evaluation and Resource Management at the University of York.

ABSTRACT

Diversity in the provision of the key public service of secondary education through different types of school is a feature of the English educational system. Much of this diversity has emanated from the distinctively different roles which Local Education Authorities have played in the development of the type of schooling for their areas. The 1944 Education Act described the resultant educational system as a 'national service locally delivered'. However, one aspect of this diversity, the continuing existence of grammar schools selecting an 'able' minority of pupils, remains a contested issue. Whereas over ninety percent of secondary age pupils are educated in 'comprehensive' schools, a few areas have retained selection of 'more able' pupils for grammar school education - along with its concomitant allocation of the others to secondary modern schools.

Selective grammar schools have typically been given a high ranking in published league tables of schools' examination results, a fact which has led to claims that these schools offer the best education for pupils. The emergence of statistical techniques making more sophisticated comparisons between such schools and others has been the motivation of the present paper. It compares the performance of able pupils in grammar and other types of school using value-added techniques on a pupil-level basis using recently available national datasets. In addition it considers the performance of 'selective systems' of educational provision for *all the pupils* in selective areas, compared with what occurs in fully comprehensive systems of educational provision.

The paper finds no evidence for the superiority of either grammar schools nor selective systems of educational provision; indeed any advantages appear to lie with those schools and systems organised on non-selective lines.

1. INTRODUCTION AND BACKGROUND

Concern about the 'optimal' form of educational provision for secondary schools in each local area has been apparent since well before 1965 when the Department of Education instructed local authorities to draw up plans for the implementation of 'comprehensive' reorganisation of secondary education (DES, Circular 10/65). For many years before this, there had been concern at the possible lack of equity in the selection process itself, and about the inadvisability of making 'once-for-all' decisions about children's educational future direction at the age of ten or eleven years. The Labour government, elected in 1964, made this issue one of its priorities for action.

Prior to the issuing of Circular 10/65 the organisation of schools locally was the preserve of the local education authority, subject to general provisions of the 1944 Education Act. The 1944 'Butler' Act encouraged the setting up of a 'tri-partite' system comprising grammar, technical and secondary modern schools. However, in the period between 1944 and 1965 considerable local diversity became apparent. Moves towards 'comprehensive' school re-organisation appeared in a number of places before 1965 - perhaps the most famous being the 'Leicestershire Plan' in the 1950's and the purpose built new 'comprehensive' schools in London. Interestingly, the London School Plan of 1947 specifically rejected the idea of selection and planned for sixty-four comprehensive schools, although this development was not completed until 1977.

In other parts of the country this form of re-organisation was seen as a natural development, avoiding problems associated with some 'bright pupils' not being selected for grammar schools, whilst other pupils whose later educational trajectory did not match their apparent potential. Both of these problems were seen to represent a waste of educational talent and resources. A series of government reports had drawn attention to the waste of underdeveloped talent that this represented (*see* Mortimore and Mortimore, 1986). Added to this was the evident sense of 'failure' experienced by many 'non-selected' pupils (and their parents) which cast a shadow over their later attitudes and

motivation in the secondary modern schools to which they were then, normally, consigned. 'Secondary modern' and 'second-rate' were all too easily an allusion that could be made; more particularly when the funding, the buildings and the qualifications of staff were compared.

Thus, moves towards comprehensive educational provision were already in place in different parts of the country well before the publication of Circular 10/65. Subsequently, re-organisation took off fairly rapidly; with the most rapid expansion (and consequently the largest 'destruction' of grammar schools) taking place in the years of the Conservative government from 1970 to 1974 under the guidance of Margaret Thatcher, then Secretary of State for Education.

Table 1: the change in character of secondary schools from 1965 to 1995

School Type	1965	1975	1985	1995
Comprehensive	4%	57%	87%	90%
Grammar	22%	12%	5%	4%
Sec Modern	64%	27%	8%	6%
Other	10%	4%	1%	~

(Source: DES/DFE Statistics of Education Vol. 1 for selected years)

Progress, however, was not uniform. Certain parts of the country resisted these moves, retaining selective educational systems either in part, or over complete local authority areas. By the early 1980's, however, many local authorities had fully implemented comprehensive secondary education, though grammar (and secondary modern) school systems still remained in some areas. This situation provided a basis for comparisons to be made between 'comprehensive' and 'grammar' schools, and, more specifically, about the relative effectiveness of these different forms of educational provision. It is this debate which has been re-ignited in the late 90's by the in-coming Labour government's

promise to allow parents to decide 'whether grammar schools should continue to select pupils by academic ability'. (DfEE, 1999).

2. CONTROVERSY IN THE EARLY 1980's

Concern about alleged 'under-performance' of comprehensive schools made its appearance in debates in the House of Commons (where a Conservative government was in office) after the passing of the 1980 Education Act which made the publication of secondary schools' examination results mandatory. The enhanced public dimension now attaching to the issue of comparative school performance saw a number of attempts to make such comparisons public. The questions which have subsequently arisen about what the 'evidence' leads one to conclude can be traced to a concern that comparisons between the outcomes of pupils in different schools be firmly and unequivocally based on appropriate assumptions. The degree to which the various contributors to this debate have made different assumptions has, in some measure, been responsible for the continued sharp differences of opinion which still occur between proponents of differing views.

A number of major contributions to the development of this debate took place within this period. First was a research report commissioned by the Department for Education and Science to evaluate aspects of progress in selective and non-selective schools in England (Steedman, 1980). This study used data from the National Child Development Study of individuals born in a given week of March 1958. The data used related to pupils' experiences and educational performance made available from the 1974 sweep, when its members were sixteen years old. The study found that the progress which pupils made in comprehensive and selective schools was not significantly different from each other once allowance had been made for the relatively lower entry attainments and social class background of pupils who went to comprehensive schools.

The findings of the NCDS study were subjected to hostile criticism from the Centre for Policy Studies (Cox et al, 1980). This criticism reinforced the increasingly 'political' polarisation of the debate - with its implication that those who challenged the assumption of the 'superiority' of selective education were biased, whilst those who supported it were not.

Rutter's influential study of the twelve London comprehensive schools (*Fifteen Thousand Hours*; Rutter et al, 1979) - which emphasised the diversity of practice and outcomes within them - did not address the issue of comparisons between school *systems*. His work did nevertheless raise important questions about the sources of variability in the outcomes which schools achieved. Rutter's work re-established the view, somewhat contested during the 1960's and 1970's, that 'schools *do* make a difference', and that performance is not pre-determined by pupils' socio-economic or other characteristics.

Directly addressing the question of whether 'areas with grammar schools' made a difference by obtaining better examination results than those which did not, was the issue tackled in '*Standards in English Schools*' (Marks et al, 1983; 1985). These studies claimed to provide evidence that *areas, which had retained selection, obtained better results than those which had not*. Similar ideas were advanced in Lord's pamphlet '*Value for Money in Education*' (Lord, 1984). In the earlier study Marks et al used data from some fifty-four English LEAs, grouped into three categories using census data relating to socio-economic disadvantage. They reported that pupils in areas which had retained selection performed better than those with comprehensive forms of organisation. In a subsequent study (Marks et al, 1985), the authors refined their methodology by incorporating a measure of LEAs' socio-economic advantage, one finding of this study was that areas which had retained selection still achieved better results on one performance measure (the average number of higher grade O-levels/CSE grade 1 passes). However, when a more 'useful measure....(of) the attainment of the whole ability range taking GCE and CSE examinations' (the points score which credits pupils with

7 points for an A grade, 6 for B and so on) was utilised they found no difference at all between the performance of the two systems. (Marks et al, 1985; *Table 13 column 3, p19*). Curiously, subsequent comments on the implications of the results appeared to ignore this finding.

The conclusions reported in the first of these studies were strongly contested by Clifford and Heath (1984) and by Gray et al (1984, 1989) who argued that the authors' method of 'controlling' for social class did not adequately reflect the differences in intakes of schools within LEAs. When Gray et al (1984) used the same data sources as those available to Marks et al (1983), they reached substantially different conclusions. The Department of Education and Science meanwhile published two Statistical Bulletins (DES, *SB 16/83 and SB 13/84*) which, for the first time, provided a common *socio-economic* framework within which to evaluate the outcomes of local authority education performance. These gave some support for the contention that areas which had retained some degree of 'selection' education produced better result. However, the methodology used for 'taking account' of differences in schools' populations was questioned and the actual level of 'advantage' for selective systems was found to be 'very small' in Gray et al.'s (1989) critique of these and other studies.

One issue of considerable concern that arose at this time was that the 'league tables' produced by these Bulletins (and further published interpretations of the data they supplied) were highly volatile. This volatility raised considerable doubts that the methods then available could answer the critical questions with any degree of authority. Gray and Jesson (1987) showed that by slightly altering the assumptions the ranking of local authority performance could vary very substantially. They carried out eight ranking exercises using available data, and in each of these 'evaluations' of LEA performance only one LEA (Harrow, a fully comprehensive Outer London Borough) retained its position in the 'top ten' LEAs throughout all of these ranking procedures. Another LEA, Bromley (an Outer London LEA, but one which had retained some selection), moved from ninth in one table (given in a Ministerial response to a question tabled in the House of Commons) to 'bottom of the league' of 96

LEAs in two separate versions of these 'league tables'; one of which was in DES' Statistical Bulletin 13/84.

A number of other studies, summarised in *The Grammar School Question* (Crook et al., 1999) also gave little support to the strong conclusions that Marks et al. had claimed in their studies.

3. IDENTIFYING TYPES OF LEA SCHOOL ORGANISATION/PROVISION

The data used by the two Statistical Bulletins raised important questions about whether *any* 'characterisation' of local authority secondary school provision (for example by making contrasts between those which had '*retained some selection*' and those which had not) was an appropriate way to proceed in pursuing the debate about the merits of grammar as opposed to comprehensive schools. Some LEAs had fully selective systems; others retained the odd 'isolated grammar' amongst the majority of other schools which were predominantly comprehensive; others were fully comprehensive in organisation. It was clear that the actual forms of secondary school provision, across all of the 96 LEAs then extant, were very diverse; and that characterisation of *local authorities* was likely to be very crude and imprecise except for those which were either 'fully selective' or 'fully comprehensive' in their provision. Even the *naming of schools* was no guide to their actual characteristics - numbers of schools had retained their 'grammar school' title when selection by (or into) them had been abolished. There were also concerns, expressed somewhat earlier, that so-called 'comprehensive' and 'grammar' schools could not exist side by side - and if schools were called '*comprehensive*' in such situations they were much more likely to be '*nothing but misnamed secondary modern schools*' (Rhodes Boyson, 1969).

Since the late 1980's there have been relatively few new studies that have addressed the issue of the comparison of selective and comprehensive systems of education, although Marks et al (1986, 1991) did publish further accounts of schools in ILEA and a wide-ranging critique of English education

generally. This latter study regretted the absence of any specially commissioned government study of relative pupil performance in comprehensive and selective schools, a concern the present author fully shares. Of particular interest for later developments was Marks' criticism that:

'the elevation of social class into the central position in the debate is mistaken, The crucial questions to ask about pupils and about their education are.....whether their abilities and aptitudes are being recognised and encouraged'

(Marks, 1991)

However, whilst most of the studies cited above indicated that there was little evidence for the superiority of one system over the other (using forms of data then available) they also indicated that there was a much greater need to focus on the performance of the *individual pupils* who passed through the systems, and on the *progress* which they made from their *prior attainments*. To do so would be in sharp contrast to previous practice, which had focused mainly on aggregate levels of performance within LEAs or schools, and on the aggregated socio-economic characteristics of these.

4. THE EMERGENCE OF PUPIL-BASED VALUE-ADDED METHODOLOGY

Subsequent work in this field has been based almost entirely on pupil-level data, and more particularly on data which references pupils' educational as opposed to their socio-economic characteristics. Such data, it is now widely agreed, provide the only firm basis for adequate (statistical or even common-sense) comparisons between schools to be made. The seminal contributions of Aitkin & Longford (1986) were based on data supplied by Gray et al. who had explored these issues through the medium of the ESRC funded *Contexts Project* (Gray et al., 1986) The methodology to do this developed rapidly in the mid to late 1980's, most notably through the contributions of Goldstein (1987); Bryk and Raudenbush (1992) as well as further developments from Aitkin and Longford themselves. As a consequence, there emerged a distinctive school of 'Value-Added' or 'school effectiveness' research which has dominated the study of comparative educational performance from around this date

onward. One particularly relevant paper of this period, published by Gray et al., entitled '*The search for a fairer way of comparing schools' examination results*' (Gray et al.,1986), set out the authors' view of the *necessary conditions for making appropriate comparisons between schools' performance*. In summary these were:

- **A:** First, that appropriate measures of performance are used; the authors contrasted the 'threshold value' approach - using the percentage of pupils in each school obtaining, say, five or more 'good' passes, to a more comprehensive measure of each pupils' examination performance in each of the examinations they took; their recommendation was to use the points score developed by ILEA allocating seven points to an 'A' pass and one to a 'G';
- **B:** Second, that appropriate 'contextualisation' is required. They explained that 'contextualized' data must reflect pupils' abilities, attainments or socio-economic circumstances; and that strong preference be given to data which relate to the individual pupil, reflecting their prior abilities and attainments; and,
- **C:** Third, that statistical 'models' are used which are appropriate to the underlying structure of data, and in particular to its 'nested' nature; that is: using procedures taking account of the fact that pupils in given classes and particular schools share more 'in common' with each other than with 'similar' pupils in other classes and schools.

Each of these conditions can now be met to a greater or a lesser degree.

- **a.** The Performance Tables (DfEE, 1998) listed secondary schools' performance using four 'measures'; three of which had been in use previously and were all 'threshold' measures (percentage of pupils achieving 5 or more A* to C passes; 5 or more A* to G passes and at least one A* to G pass); the fourth was a new one: the GCSE/GNVQ Points score which allocated points to every pupil for every grade or pass that they achieved in each of their GCSE/GNVQ examinations. This additional measure is a truly 'comprehensive' measure of pupils' and schools'

performance - since it reflects the results of every examination passed by every pupil in each school, rather than reflecting the (undifferentiated) outcome achieved by only a minority of pupils (% 5+ A* to C passes) which has, to date, been the major source of comparisons between schools. *This meets point A made by Gray et al., (1986).*

- **b.** With the implementation of National Curriculum assessment, national data is now available on the comparative 'prior' performance of pupils at points in their educational careers prior to their GCSE/GNVQ examinations. This provides a useful, nationally available, 'prior attainment' framework both for individual pupils and also, in aggregate, for their schools. *This goes a considerable way to meet point B made by Gray et al., above.*
- **c.** Two ways of presenting information are now available - one uses sophisticated statistical methods (including regression and multi-level modelling), which for the most part provides the intellectually coherent background to any comparative evaluations of schools' performance; we utilise these methods in the penultimate section of this paper. Another, relying on 'benchmarks' derived from national data shows the ranges of outcomes achieved by pupils with given prior attainments. It is the latter which is now published in the annual DfEE/QCA/OfSTED *Autumn Package* - and which is presentable in highly visual and tabular form. This is a useful form of presentation for individual schools - since it allows them to place their own pupils' performance, for the very first time, within an intellectually coherent, national, prior attainment framework. Evaluations of performance are based on comparisons with percentile rankings of 'similar' schools and 'similar' pupils. *These go a considerable way to meeting point C made by Gray et al., above.* Additional data is now available together with sophisticated modelling techniques which offer considerable scope to re-examining the issue of selective and comprehensive forms of educational organisation.

5. THE FIRST APPLICATIONS OF VALUE-ADDED METHODS TO EVALUATE SCHOOL PERFORMANCE

The early 1990's saw a period of increased hostility by many in the teaching profession to a series of government actions designed to change the nature of the public education service in ways which appeared alien to previous practice. In particular, the outcry over the 'inappropriateness' of league tables was such that many local authorities began to evaluate their own schools' performance with some form of 'Value-Added' analyses. This movement grew rapidly in these years, and such was its impact that government set up a review, under Ron (later Sir Ron) Dearing, to evaluate the role of 'Value Added indicators of school performance'. The report was published in 1994 (SCAA, 1994) and recommended a pilot study of the application of 'Value-Added' methodology to both secondary and primary school 'terminal' examinations (at GCSE and Key Stage 2 levels). This pilot study carried out by Professor Carol Fitz-Gibbon and colleagues at the University of Newcastle-upon-Tyne, recommended use of pupils' prior attainments (as measured by their Key Stage assessments) as the 'context' within which each school's examinations are evaluated (Fitz-Gibbon, 1997).

Simple Value-Added diagrammatic frameworks have a horizontal axis, which measures the *prior attainments* of the pupils or schools; their *performance outcomes* are measured on the vertical axis. Each 'unit' (pupil; school or other entity) is plotted on this chart; this allows a line of 'best fit' to be drawn through the average levels of outcome for each level of prior attainment. In many instances this line is constrained to be straight, as shown in the diagram, but this is not an essential feature of Value-Added frameworks.

A simplified illustrative version of a Value Added framework is shown in Figure 1 below:

Value Added Performance Evaluation Framework

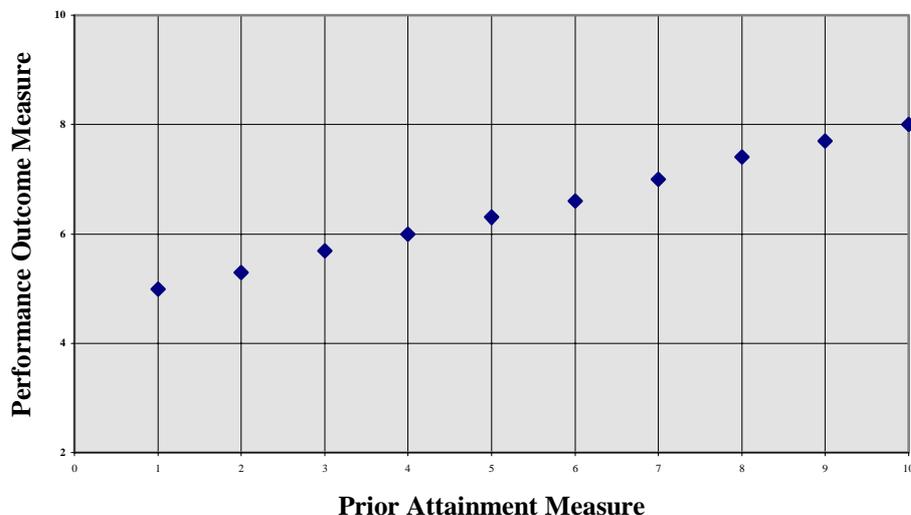


Figure 1

Evaluation of performance is carried out with reference to the line of best fit shown in Figure 1: outcomes above this line show performance which is ‘better’ than would be predicted given the prior attainment; whilst performance below the line shows performance which is ‘worse’ than would be predicted by their prior attainment. There is normally a scatter of points, each representing an individual pupil’s or school’s performance, some of these are ‘just above’ and others ‘just below’ the line. Simple statistical methods can be applied to decide ‘how far’ a point needs to be ‘above’ or ‘below’ the line to provide genuine evidence that this performance is ‘better’ or ‘worse’ than average.

A recent report *Value Added Pilot Study Supplement to Performance Tables for Secondary Schools* (DfEE, 1999) made the following observations about ‘Value-Added’ frameworks:

Value Added measures are intended to allow fairer comparison between schools with different intakes. For example, the pupils at school ‘A’, in a ‘leafy suburb’, may generally achieve above the national average at GCSE/GNVQ. In school ‘B’, in a more disadvantaged location, pupils may generally achieve below the national average at GCSE/GNVQ. But which school adds more value? Using Key Stage 3 results as a starting point, it is possible to see whether groups of pupils at each of these two schools have made more or less progress in terms of their total GCSE/GNVQ point scores.

School 'A' may well have more pupils gaining 5 A* - C passes but the pupils in school 'B' may have made more progress relative to their Key Stage 3 starting point and thus have higher 'Value Added' scores.

There are several different possible ways of measuring Value Added. All systems use performance data but some also have additional factors such as the take up of free school meals. The SCAA (School Curriculum and Assessment Authority) National Value Added project found that the best predictor of future performance is a pupil's previous attainment regardless of other factors. The most straightforward system, which is used for this pilot, and in the work which DfEE and QCA (Qualifications and Curriculum Authority, formerly SCAA) are doing on Value Added for school improvement, uses only performance data.

So the key to full Value Added measures is the availability of suitable data on the attainment of individual pupils at different stages of their education. This means that progress can be measured between one stage and another for that pupil, and, by aggregating the pupil 'scores', a Value Added measure calculated for the school. Data in a suitable form is only now becoming available. First to come on stream are the Key Stage 3 test/task results which, compared with the 1998 GCSE/GNVQ exam results, have been used to provide the pilot Value Added measures in these tables. Data for other key stages will become available over the next few years. GCSE/GNVQ results for comparison with Key Stage 2 results will be available from 2003 at the earliest.

(also available on DfEE website: http://www.dfes.gov.uk/performance/vap_98/docC.htm)

A Value Added approach has, therefore, received some official recognition as a means of overcoming some of the deficiencies of league tables in which those with favourable intakes are likely to perform particularly well.

The Technology Colleges Trust in association with DfEE commissioned two reports, one in 1996 and the other in 1998 on the use of pupils' Key Stage 3 attainments, as a basis for assessing their schools' GCSE performance. The first of these, (Jesson, 1997), showed clearly how this form of analysis could provide very powerful information about the relative 'effectiveness' of schools in delivering GCSE outcomes to their pupils. The second took this further, and showed how performance in a range of individual subject areas in each school could be appropriately compared and evaluated. (Jesson, 1999). Each school was compared to all others through its effects on its pupils using a sample of many thousands of pupils with similar characteristics within all the schools that were considered within the project.

6. EVIDENCE OF VALUE ADDED PERFORMANCE EVALUATION IN OFFICIAL PUBLICATIONS

We have quoted above from the Introduction to the DfEE's Report on the *1998 Value Added Pilot; Supplement to the Secondary School Performance Tables (DfEE, 1999)*. This report provided insights into the question that lies at the heart of this paper in respect of comparisons between performance outcomes between schools of differing types. The study it reported was of a 'broadly representative sample of 200 volunteer secondary schools... covering nearly 30,000 pupils from the national cohort of 15-year-olds'.

Of the 200 or so schools, some were 'Independent' and so were excluded from the analysis reported here. Of the remaining 186 schools, 157 were comprehensive schools; 19 were grammar schools and ten were described as 'modern' schools. The GCSE performance of three distinct groups of pupils was used to capture the effects of schools on different parts of the 'prior attainment' spectrum. Pupils were classified into one of three groups on the basis of their 'average Key Stage 3 test levels' awarded in 1996. These groupings were:

'low prior attainment' - average Key Stage 3 level 4 or below:

'average prior attainment' - average Key Stage 3 level between 4 and 5.66; and,

'high prior attainment' - average Key Stage 3 level above 5.66.

Since grammar schools select the *most able* pupils, it was appropriate to use the evaluations of performance for 'high prior attainment' pupils to carry out comparisons between *grammar and selective schools'* effects on pupils in this group compared with that in other types of school. Somewhat surprisingly, and reflecting our earlier reference to the (sometimes inappropriate) 'naming' of schools, each of the ten 'modern' schools had sufficient pupils of 'high prior attainment' to feature in this evaluation

Each school's performance with this group of pupils was reported on a five point (A to E) scale, identical to that proposed in the Department's Progress Measure described above. Table 2 shows the results ranging from Scale Point A representing 'very good' performance' to E for 'very poor' performance.

Table 2 shows the percentage of schools gaining evaluation in each of the A to E categories

Progress Measure	Comprehensive Schools	Grammar Schools	Grammar & Modern Schools
	(n=158)	(n=19)	(n=29)
A (best)	10%	0%	0%
B	23%	21%	14%
C (average)	36%	53%	52%
D	20%	26%	21%
E (worst)	10%	0%	14%

The small number of schools makes it unrealistic to apply conventional statistical tests to this data. It should be noted here that 14 out of the 19 grammar schools were in areas which had retained a high degree of selection; 9 out of the 10 modern schools were also similarly placed; the other 'modern' school was in Ripon where there was also a grammar school. Thus, the percentages in the final column related, in the main, to evaluations of performance in 'selective' areas, although direct comparison between performance in comprehensive and grammar schools should be made using columns 2 and 3.

One observation can be made from information presented in Table 2: able pupils in these comprehensive schools appeared to make *at least as good progress* as their counterparts in grammar schools. *Ten percent of comprehensive schools were rated 'A' for their performance with these pupils; none of the grammar schools were so evaluated.*

Even when categories A and B were combined (as was done for the national Progress Measure), comprehensive schools still appeared to perform well in comparison with grammar schools: one *in three* outcomes was assessed within this range against only *one in five* of the grammar schools.

Using the comparisons available from the final column (which related, in the main, to 'selective systems' of schooling), the comparison indicated rather more strongly the relative strengths of comprehensive schools: one in three comprehensive schools was in these 'top' two categories against only one in seven of the schools in 'selective systems' of LEA provision. It is true that the evidence available here relates only to a small number of schools - those involved in the DfEE Pilot exercise - and that general conclusions cannot be drawn from this. However, the indications did not offer support for claims that these grammar schools (and their selective systems of education) 'did better' than others, even in respect of the performance of the most able pupils. This represents a challenge to the claims of advocates of selective education that grammar schools provide better performance.

7. GCSE PERFORMANCE ASSESSMENTS PUBLISHED IN 1998 GCSE/GNVQ PERFORMANCE TABLES: THE FIRST EVER NATIONAL VALUE ADDED EVALUATION OF SECONDARY SCHOOL PERFORMANCE

Prior to the publication of the 1998 Performance Tables, statisticians at DfEE devised a method of translating Professor Fitz-Gibbons' recommendations into practice. The structure of this is described on the DfEE Standards Internet Site:

Schools were placed in one of twenty-seven narrow bands of KS3 average performance. The difference between each school's GCSE/GNVQ average point score and the median of its band was calculated and schools were ranked across the whole spectrum of KS3 performance. Schools were then placed in one of five groups identified by their position in respect of the 95th/75th/25th/5th percentiles. An individual school's progress measure band can be contrasted with its position in the appropriate benchmark table. Because the calculation of progress measures involves comparing relative GCSE/GNVQ achievements within much finer KS3 bandwidths, it may be the case that a school may have a higher, or lower, benchmark position than progress measure band.

(<http://www.standards.dfee.gov.uk/library/publications/autpack/gcse/section3/h>)

Thus each secondary school in England had its GCSE/GNVQ performance evaluated on a common basis: *the first time that this had ever been possible*. That the evaluation did not cover the *whole* of the years of secondary education was, of course, a disadvantage, but the nature of the National Curriculum assessment framework, and its implementation timetable, guaranteed that such evaluations would not be available until the year 2003 at the earliest. So, to repeat: this evaluation was *the first ever Value Added analysis of the GCSE/GNVQ performance of every secondary school in England*. As such, its findings were and are of considerable significance for the re-activated debate about the effectiveness of differing types of school and schooling systems.

The 1998 DfEE Performance Tables contained a tick (✓) for each school whose performance was adjudged to be in the 'top' twenty-five percent of schools serving 'similar' pupils. Since there are around 3200 secondary schools in England, somewhere around 800 schools received this positive evaluation.

The distribution of these positive evaluations was by no means evenly distributed across different groupings of schools nor across the LEAs which were responsible for them. Of considerable interest, however, was the performance of grammar schools in the ten LEAs identified by DfEE as those where parental ballots would need to take place to decide: *'Whether grammar schools should continue to select pupils by academic ability'* (extract from DfEE publication: *Grammar School Petitions and Ballots*).

There are ten areas where schools are covered by such local education authority ballots, they are: Bexley, Buckinghamshire, Kent, Lincolnshire, Medway Towns, Slough, Southend, Sutton, Torbay and Trafford. Together these LEAs contain 94 of the existing 164 grammar schools. The other 70 are found within other, mainly comprehensive, LEAs. Table 3 below shows how well these grammar schools (and their associated secondary moderns) fared.

Table 3 shows the GCSE performance evaluations of all schools in ten ‘selective’ LEAs compared with evaluations of the rest of the nation’s schools

LEA	Total No of Schools	No of Grammars	No of Sch with √	Percentage with √	No of Grammar √	Percentage with √
Bexley	16	4	7	44%	0	0%
Bucks	36	13	8	22%	4	31%
Kent	105	33	14	13%	5	15%
Lincs	64	15	8	13%	1	7%
Medway	20	6	3	15%	0	0%
Slough	11	4	4	36%	2	50%
Southend	14	4	1	7%	0	0%
Sutton	14	5	3	21%	1	20%
Torbay	8	3	0	0%	0	0%
Trafford	18	7	0	0%	0	0%
All Select Areas	306	94	48	16%	13	14%
All Other	app 2800	app 70	app 750	app 27%		

(Source: DfEE Secondary School Performance Tables, 1998)

A Chi-squared test applied to the summary data gave $X^2 = 18.1$ (with 1 degree of freedom), rejecting the hypothesis that selective systems provided the same level of positive evaluation as comprehensives. Two observations arose from inspection of the information in Table 3:

A . Overall, the last two lines showed that ‘selective’ areas had fewer ‘good performing’ schools than areas which had other forms (mainly comprehensive) of organisation. *Only around 16 percent of schools in selective areas ‘did well’; compared with over 27% in mainly comprehensive, areas.*

B. Taking grammar schools on their own, only 13 out of the 94 in these ten LEAs were classified in the ‘top’ category of performance’: *this is just over 14% - about half the level which would be expected if grammar schools were providing as good performance as other types of schools.*

Thus, the first ‘official’ Value Added evaluation of GCSE performance in English schools, even in the somewhat limited form in which it was published, provided two indications:

- **a.** that ‘*selective school systems*’ did not ‘do better’ for their pupils than did ‘comprehensive education systems’ as a whole; and,
- **b.** that individual grammar schools were only half as much in evidence amongst ‘good performers’ than would be expected if they were performing as well as comprehensive schools.

These indications offered a further challenge to the validity of claims that grammar schools provided ‘superior performance’ for pupils overall when compared to comprehensive schools. To these should be added the observation made in respect of Table 2 - that there was no evidence from the DfEE 1998 Value-Added Pilot Study that grammar schools or selective systems as a whole ‘did better’ for more able pupils either.

The data which we have utilised up to this point has been derived from published sources (as indicated) and represents substantial aggregation of the individual pupil-level information from which it was derived. In the next section of this paper we have utilised individual-level data relating to most of the pupils in state-maintained schools who took their GCSE examinations in 1998. The matched data comprised about 550,000 individual records comprising pupils’ Key Stage 3 performance in 1996 and their GCSE outcomes across all subject areas in 1998 within most of the maintained secondary

schools in England. This data was made available by courtesy of the Data Access Board of the Department for Education & Employment to whom grateful thanks are extended. The analysis and conclusions, however, remain the sole responsibility of the author.

8. USING NATIONAL DATA ON ‘PRIOR ATTAINMENT’ AND GCSE OUTCOMES

The data sources identified pupils’ Key Stage 3 performance (in most cases in each of the three foundation subjects of English, Mathematics and Science), their GCSE outcomes (in individual subject areas as well as the aggregate points achieved), the database also identified pupils’ gender and the DfEE establishment number (or school) within which they took their GCSE examinations.

The 1998 Form 7 specification identifying the ‘status’ of each school was utilised in summary form to construct a variable identifying those schools that were either grammar schools or were not. This information was matched to the pupil-level data file that comprised a single line record of their ‘prior attainment’; their GCSE performance and other personal and school-based factors. Data was unavailable for a few schools, and problems with data collection have continued to arise subsequently. Even so, the availability of relevant information for most of the nations’ pupils and schools is one of the successes of the National Data Collection exercise conducted by QCA for the DfEE. Matched information on just over 500,000 pupils formed the basis of the analyses which follow. We excluded data from all Independent and Special schools and also from all pupils whose Key Stage 3 record comprised less than two assessments. The database allowed classification of the educational structure of English maintained secondary schools to be identified. Four ‘types’ of school were identified as reported in Table 4 below.

Table 4 showing the numbers of pupils and schools in each of four ‘types’:

School type	Pupil No’s.	Percentage	School No’s	Percentage
Comprehensive	448725	89.4%	2714	87.1%
Sec Modern (Selective LEAs)	34370	6.9%	246	7.9%
‘Isolated’ Grammars	5696	1.1%	54	1.7%
Grammar (Selective LEAs)	12895	2.6%	101	3.2%

It should be noted that ‘Secondary Modern’ schools included all non-grammar schools in those areas where selection of around twenty percent or more of the pupil population for grammar schools was in place. ‘Isolated’ Grammar schools exist in certain areas of individual local authorities, for example, North Yorkshire has 42 maintained schools: three of which are grammar schools; two in Skipton and one in Ripon.

We were able to use pupils’ prior attainments to derive the composition of each school type. The overall prior attainments of all pupils were ‘banded’ into five groups of approximately equal size. For convenience we labelled each group as follows:

Group 1: ‘high prior attainment’ (pupils with average Key Stage 3 levels > 5.67);

Group 2: ‘above average prior attainment’ (average Key Stage 3 levels between 5.2 and 5.67);

Group 3: ‘average prior attainment’ (average Key Stage 3 levels between 4.5 and 5.2);

Group 4: ‘below average prior attainment’; (average Key Stage 3 levels between 3.67 and 4.5);

and **Group 5: ‘low prior attainment’** (Average Key Stage 3 levels < 3.67)

Table 5 shows the distribution of pupils across prior attainment groups in each type of school.

School type ↓	Prior Attainment⇒	High	> Avge	Avge	< Avge	Low	School Key Stage 3	
							Mean	Std Dev
Comprehensive		16.6	19.9	22.8	22.9	17.9	4.74	0.454
Secondary Modern		8.5	19.7	26.8	26.8	19.2	4.57	0.373
‘Isolated’ Grammar		82.1	13.2	3.1	1.1	0.5	6.43	0.453
Grammar (Selective LEAs)		78.9	18.5	2.4	0.1	0	6.31	0.245
Overall		18.4%	19.7%	22.3%	22.3%	17.3%	4.81	0.443

Table 5 shows that substantial differences between ‘school types’ occurred both in the average Key Stage levels of their pupils, and in the distribution of pupils’ prior attainments within each type. It is interesting to note the much lower standard deviation of pupils’ prior attainments in selective LEAs - representing the choices made in those areas to pre-select pupils for particular forms of educational provision, and creating more homogeneous within school groupings. The variability in prior attainments amongst ‘Isolated’ Grammar schools is comparable to that of Comprehensive schools, suggesting that local conditions of entry into such schools may be much more variable than those applying to Grammar schools in selective LEAs; this also suggests that these schools do not form as clear a ‘typology’ as do Grammar schools in selective LEAs. ‘Isolated’ Grammar schools form a very small minority of the schools within their LEAs, which are, therefore, best described as ‘predominantly comprehensive’. The mean Key Stage 3 level of schools in ‘selective systems’ of LEAs was 5.04, substantially higher than that of those ‘with comprehensive systems’.

The percentage of ‘high prior attainment’ pupils in Comprehensive schools, whilst much lower than those in the relatively small number of Grammar schools, nevertheless indicated (using data from Table 4) that the great majority of such pupils (around 74,500 out of 92,300) were educated in Comprehensive schools. Their comparative performance evaluation will be addressed in a later paper.

Tables 4 and 5 above give an overview of the educational ‘provision’ and pupil context of the 3000+ secondary schools that provide secondary education for the majority of pupils educated within the state-maintained sector. It provides one useful overview of the system in ways that have not been possible heretofore; it does not, however, represent a full or complete account insofar as it is based on information relating to progress in the final two years of secondary education. Neither does it provide any account of the differing socio-economic contexts within which these schools exist. However, for present purposes the data does provide a further means of exploring the differences in outcomes that the different ‘systems’ of secondary school organisation produce.

In passing, it is interesting to note that proponents of ‘selective’ education have often stated that ‘Secondary Modern schools provide a very effective complementary education for pupils not selected for Grammar school’, which identifies the need to evaluate the impact of the system of provision on *all* pupils, not simply the performance of the minority who attended grammar schools.

In the Sections that follow our emphasis will be on contrasts between *non-selective and selective* systems of educational provision, although we will also continue to look at outcomes in each of the four types of school identified in Tables 4 and 5 . We have used information relating to Comprehensive schools and contrasted this with that from the combination of Grammar and Secondary Modern schools. This allows contrasts to be drawn directly between outcomes from the two types of system. This information should provide a useful contribution to the debate on forms of provision and their outcomes; restricting the debate only to the performance of Grammar schools would ignore the important majority of pupils in ‘selective’ areas who are educated in Secondary Modern schools.

9. CHARACTERISTICS OF ‘SELECTIVE’ LEAS

We calculated the proportions of pupils selected for grammar school education in each of the 150 LEAs. This identified fourteen LEAs where ‘substantial’ selection of pupils for grammar schools (and their counterpart secondary modern schools) took place. The criterion used for defining an LEA as ‘selective’ was that around twenty per cent or more of its Year 11 cohort of pupils were in grammar schools. Table 6 shows the percentages for these LEAs in 1998, and their pupils’ average KS3 level.

Table 6 shows the percentage of pupils selected for grammar schools (and also the percentage in Secondary Moderns) in LEAs which had retained such selection to a substantial degree.

LEA	Grammar School	Secondary Modern	Average KS3
	Percentage of pupils		
Bexley	22%	78%	4.93
Sutton	31%	69%	5.47
Wirral	22%	78%	4.92
Trafford	30%	70%	5.06
Buckinghamshire	40%	60%	5.29
Poole	20%	80%	5.26
Bournemouth	19%	81%	5.12
Reading	23%	77%	5.04
Slough	36%	64%	4.82
Torbay	29%	71%	5.33
Southend	30%	70%	5.19
Kent	29%	71%	5.05
Rochester & Gillingham	25%	75%	4.92
Lincolnshire	20%	80%	4.99
All LEAs	3.7%	6.9%	4.80

The final column of Table 6 shows that pupils in ‘Selective’ LEAs had *generally* higher average prior attainments than in the country as a whole. This places even greater emphasis to the need to evaluate school and pupil performance by Value Added means, taking account of pupils’ prior attainments.

10. OUTCOMES FOR PUPILS IN COMPREHENSIVE AND SELECTIVE SCHOOLS

In Table 7, below, we show the GCSE outcomes which pupils achieved in the two contrasting types of educational provision and in each of the prior attainment groups we have identified above.

Table 7 showing the average GCSE performance for pupils in each prior attainment group in Comprehensive and Selective schools

Prior Attainment Group	Comprehensive	Selective LEAs	Comprehensive	Selective LEAs
	GCSE Points Score (s.d. in brackets)		% achieving 5+ A* to C passes	
Low	12.9 (10.6)	12.6 (9.9)	1.0 (0.07)	0 (0.05)
Below Average	24.7 (13.0)	23.8 (12.4)	8.1 (0.27)	6.0 (0.27)
Average	36.0 (13.3)	34.4 (13.1)	39.1 (0.49)	35.4 (0.48)
Above Average	46.1 (12.8)	44.9 (12.7)	78.7 (0.41)	76.5 (0.42)
High	57.9 (15.0)	58.4 (14.4)	94.3 (0.23)	95.1 (0.22)

The average performance of pupils in Comprehensive school systems was higher than that for pupils in ‘Selective’ LEA schools for all but the most able pupils. Given the higher average prior attainments of pupils, generally, in ‘selective’ LEAs this one difference provides no evidence of superior performance for the selective system overall. The overall message of Table 7 is that pupils generally ‘did better’ in Comprehensive as opposed to Selective LEAs once account had been taken of their

prior attainments. This is essentially a ‘Value-Added’ evaluation of performance. It is now possible carry out this evaluation using data from each of the individual pupils, rather than *grouping* them as has been done in Table 7.

In order to explore this issue further we carried out a modelling exercise whereby we evaluated how well individual pupils performed in their GCSE outcomes given their:

Gender; prior attainment; and, whether they were in comprehensive or selective LEAs.

This involved use of multi-level regression analysis; first estimating GCSE outcomes for individual pupils in each type of LEA within which each school was placed; and then, modifying this to explore what, if any, difference emerged when Comprehensive and Selective systems of school organisation were compared using a single categoric identifier. The results are reported in Table 8 below.

Table 8 giving details of the progressive modelling of GCSE performance in Comprehensive and Selective systems of educational organisation.

	Comprehensive Systems	Selective Systems
Before modelling GCSE Points Score Performance		
1. Variance between Schools	74.1	182.8
2. Variance between pupils	331.5	249.8
3. % variance between schools	18.3%	42.3%
4. After modelling GCSE Points Score Performance: using pupils' Average Key Stage 3 level and		
5. Variance between schools	13.4	12.5
6. Variance between pupils	143.2	137.5
7. % variance between schools	8.6%	8.3%
8. R ² % variance accounted for	61.4%	65.3%
9. Predicted GCSE Points for a Female of average KS3 level	$37.1 + 14.4 \times (\text{AvKS3} - 4.8)$ - 3.1 GCSE Pts (if male)	$35.8 + 14.1 \times (\text{AvKS3} - 4.8)$ - 2.4 GCSE Pts (if male)
10.	For both systems combined in a single model	
11.	Before modelling	After modelling
12. Variance between schools	87.0	13.3
13. Variance between pupils	327.7	142.7
14. % variance between schools	20.9%	8.5%
15. R ² % variance accounted for	62.4%	
16. Predicted GCSE Pts for female of average KS3 level in Comprehensive system	$37.1 + 14.4 \times (\text{AvKS3} - 4.8) - 3.0$ (if male) - 1.1 (if in a 'selective' system)	
17. Std errors for model coefficients	37.1 (0.09) 14.4 (0.04)	-3.0 (0.08) -1.1 (0.25)

Table 8 provides a wealth of information about pupils' and schools within the contrasting systems that form the focus of this paper. By using pupils' prior attainment, as recommended in the report on the

National Value Added Feasibility Study (FitzGibbon, 1997) alongside their gender, almost two thirds of the variability in pupils' GCSE outcomes was accounted for ($R^2 = 62.4\%$) In consequence, the apparent very wide degree of variability in schools' outcomes (measured by the percentage of variance 'between schools' and shown in line 3 of Table 8) was very substantially reduced (see line 7 of Table 8). This indicates that although schools still varied quite substantially in their effectiveness, the degree to which they did so was very much lower than 'league-table' rankings showed - mainly because the model has taken account of the extent to which different schools educated children with differing levels of prior attainment. The remaining 'variance between schools' was around eight percent, and interestingly, was at very similar levels amongst both comprehensive and selective systems of schools. It is the 'variance between schools' that allows estimation of the Value Added measure of each school's GCSE performance.

This shows the importance of Value-Added evaluations of performance - particularly when contrasting the situation in row 3 (for selective systems) which shows, as expected, very large initial variations between schools' performances, with that in row 7. Value-Added methods of 'accounting for' performance play a major role in bringing out the essential contrasts of importance rather than leaving them embedded in 'unadjusted raw league-tables'.

11. IMPLICATIONS FOR COMPARISON OF PUPILS' PERFORMANCE IN TWO CONTRASTING SCHOOL SYSTEMS

The main conclusion about pupils' performance, however, comes from the information in the penultimate row of Table 8. The final '*model*' of GCSE performance predicted a GCSE Points score of:

37.1 for a female of Average Key Stage 3 Prior attainment (Key Stage 3 level - 4.8). For a male this is **decreased** by 3.0 GCSE Points; For pupils with other Key Stage 3 Prior attainments this score is

adjusted by 14.4 GCSE Points for each full level above 4.8 (add this amount) or below 4.8 (subtract this amount) and proportionately.

Finally, if pupils were in schools within ‘selective’ systems this estimated GCSE performance would be **decreased** by a further 1.1 GCSE Points. Thus the model of pupil performance indicates that pupils in selective systems of school organisation performed slightly less well than their peers in comprehensive systems.

A further analysis, using the same framework, but using pupils’ performance in achieving five or more A* to C passes in their GCSE/GNVQ examinations, was also carried out (details not supplied). This came to very similar conclusions, indicating that pupils in ‘selective’ systems achieved around *two percentage points lower* on this measure than their peers in comprehensive systems.

12. IMPLICATIONS FOR COMPARISON OF SCHOOLS’ PERFORMANCE IN TWO CONTRASTING SCHOOL SYSTEMS

The procedures used to derive Table 8 can also provide details of the ‘Value-Added’ measure of each school’s performance. The actual GCSE performance of each school’s pupils’ is compared with what they would have achieved had they performed in line with the ‘national’ average for pupils like them. The differences in outcome arising from these comparisons are then summed for each school to give a Value-Added sum which can be turned into a statistical indicator of whether this performance is significantly ‘better’ or ‘worse’ than expected. Normal statistical criteria can then be used to classify each school’s ‘Value-Added’ performance.

Using the resulting ‘Value-Added’ measures allowed schools’ performance to be classified into one of five Value Added categories, using the statistical criteria shown: (‘z’ is a standardised normally distributed deviate used for assessing statistical significance.) The five categories are:

A: very much better than expected ($z > 3.5$)

B: better than expected ($1.96 < z < 3.5$)

C: as well as expected ($-1.96 < z < 1.96$)

D: worse than expected ($-3.5 < z < -1.96$)

E: very much worse than expected ($z < -3.5$)

Other choices were, of course, possible for these bands; these were chosen to give a reasonably 'balanced' description of schools' performance using conventional statistical practice for the definition of category C. When schools' GCSE performance was compared in this way, Table 9 shows the outcome for comprehensive systems as opposed to selective systems of school organisation.

Table 9 showing the percentage of schools in Comprehensive and Selective systems of school organisation in each of five categories of GCSE performance evaluation

LEA Organisation	A	B	C	D	E
Value Added Evaluation ⇒	v much better	Better	as expected	worse	v. much worse
Comprehensive	4.8%	8.4%	76.1%	8.0%	2.7%
Selective	0.9%	5.2%	77.5%	11.0%	5.5%

Table 9 shows that for the great majority of schools performance was 'as good as expected'. There was little difference between the two sectors in the percentages of schools assessed in category C. However, inspection of categories A and B indicates a clear advantage for schools in LEAs where the organisation is comprehensive in nature. Some thirteen percent of comprehensive schools were evaluated as performing either 'better' or 'very much better' than expected. This was over twice the percentage for schools in selectively organised LEAs (which included both grammar and secondary modern schools). Category A evaluation ('very much better than expected') was even more weighted

towards schools in comprehensively organised LEAs; around five times as many of these schools received an A rating compared to those schools in selectively organised LEAs.

Finally an evaluation was made of the comparative Value-Added performance for each of the types of school identified in Section 13 above. These results are shown in Table 10.

Table 10 showing the percentage of each type of school in each Value-Added category

School	Perform	A	B	C	D	E
type ↓	Evaln ⇒	v much better	Better	as expected	worse	v. much worse
Comprehensive Sch		5%	8%	75%	8%	3%
Secondary Modern		1%	4%	76%	12%	7%
Isolated Grammar		0%	15%	81%	2%	2%
All Grammar Sch		0%	10%	82%	6%	2%
Selective systems		1%	5%	77%	11%	6%

Table 10 provides an overview of ‘Value-Added’ performance amongst schools that are organised in differing ways. If comprehensive schools are taken as the base for comparison, it is clear that they outperformed all other types of school as far as obtaining results that were ‘very much better than expected’. All other types of school did less well in this respect. The small and somewhat diverse group of ‘isolated’ grammar schools did relatively well when categories A and B were combined – some 15% of these schools obtained results which were ‘better than expected’ – which is very comparable to but at a marginally higher than the level achieved by comprehensive schools. However, when *all* grammar schools as a single group were compared with performance in comprehensive schools, it was comprehensive schools (with 13% in categories A and B) that outperformed the 10% of all grammar schools whose Value-Added measure of performance placed them within these categories.

Thus, as far as ‘excellence’ in GCSE examination performance is concerned it is evident that comprehensive schools had a clear ‘edge’ in this first national evaluation.

Turning now to performance which was ‘worse or very much worse than expected’, there were greater proportions of comprehensive schools in categories D and E than any of the combinations of grammar schools on their own. However, secondary modern schools did even worse in this respect. Whilst some 11% of comprehensive schools appeared in these categories, not far short of double this percentage of Secondary Modern schools were evaluated in categories D and E (19.1%).

From Table 9 the evaluation of comprehensive systems of schooling shows that 11% of these schools were assessed in categories D and E, but over 16% of schools in selective LEAs were also so assessed. So, in terms of Value Added assessments of performance the indications are quite clearly that **fewer schools, proportionately, in areas with comprehensive forms of educational organisation ‘did worse than expected’ than in areas which have retained selection to a substantial degree.**

The main thrust of this paper has been about contrasts between ‘systems’ of secondary school organisation, for which Table 9 carries the most significant message. The degree to which the information in Table 10 alerts us to the poor performance of Secondary Modern schools should not distract attention from the fact that Secondary Modern schools are an essential part of *selective systems* of school organisation. They are an integral part of the relative under-performance of selective systems. We have shown that when systems are compared using Value-Added methods for GCSE performance in 1998, it is **comprehensive systems of educational organisation** which **outperformed those which had retained selection.** Comprehensive systems have proportionately *more* schools which show high levels of Value-Added and proportionately *less* schools which show low levels than selective systems.

These results thus raise a substantial challenge to the claim that ‘grammar schools produce better examination results’. This claim, often made by supporters of retaining grammar school systems has most certainly not been affirmed by the evidence assembled in this study. In almost all of the comparisons made above the evidence points in the opposite direction. Interestingly, this is line with the indication first suggested in the final section of Gray et al’s 1984 paper.

13. DISCUSSION

The importance of these results, whilst they are, no doubt, of statistical ‘significance’, must first of all be considered for the evidence they give about the *direction* of the differences between comprehensive and selective forms of educational education. Whilst much previous work on differences in performance between selective and comprehensive systems has found ‘no significant difference’; those few studies which have reported findings in favour of selective systems have generally been criticised for methodological or other weaknesses. One of these studies, (Marks, et al. 1983), when its data were re-analysed using slightly different, but no less reasonable, assumptions, questioned whether there was any ‘advantage’ for ‘leas which had retained selection’ and, further, hinted that the evidence could be read to point in the opposite direction. The same could be said of most studies of this period - which is why the ESRC ‘Contexts Project’ was an ideally timed initiative. Much of the ‘uncertainty’ and lack of robustness in these studies was linked to their use of data at different levels of aggregation and, in some cases, of questionable relevance to the issues involved. The development of school effectiveness research, and in particular, the use of ‘multi-level’ models for exploring value-added issues, can be seen as one response to the unsatisfactory state of affairs represented by many of the studies in this field up to the mid-1980’s. The key requirement which emerged from the development of these studies was the need for data to be available on the individual pupils in each of the schools whose performance was to be compared. The present study is the first

that has been able to use individual pupil level data for most of the nation's schools, and to place it in a value-added context.

We have outlined in Section 4 above, one view of the qualities of data and methods required before more robust evaluations of the comparative performance of schooling systems could be made. We believe that many of the requirements for a more realistic overview of these systems have now been met. Given the improvement in the quality and consistency of the data which is now available (all pupils' prior attainments have been measured within the same National Curriculum test framework; their GCSE outcomes arise from examinations which are again, subject to national standards), it is important to stress the consistency in the findings reported from each of the sources of information utilised in this paper. They almost all point in the same general direction.

Each part of the evidence base assembled here carries a consistent message – that selective school systems, and pupils in selective systems, do not perform at higher levels than those in comprehensive systems. This is a major finding, and, whilst it would be unwise to put precise figures on the degree of advantage which pupils and schools in comprehensive systems have over those in selective systems, it is to be hoped that more detailed analysis of further years' of results will lead to greater understanding of this issue. Some estimates have been included above, but further work needs to be done in this area, preferably using data collected on a similar basis and over a number of years.

Whilst it would be unwise to suggest that the work reported here amounts to a 'justification' of comprehensive schooling in terms of its examination results, (there is always room for improvement within any system), it does nevertheless raise very serious questions about the *current* validity of claims that selective education systems deliver 'better results'. There is no evidence in support of this claim from any of the evidence reported here.

The present study, however, is not without its own limitations - and these must condition the weight with which its findings should be considered. First, as already acknowledged, the data utilised applies

only to the *last two years of pupils' schooling* - and although this is an inevitable consequence of the way the National Curriculum Assessment framework was implemented - it does leave open further questions about the effects of schooling systems on pupils' *entire* secondary school careers. This issue will be addressed when appropriate data becomes available, but that is not expected for a number of years yet, and, given the volatility in the nature and implementation of the early Key Stage 2 assessments, too much should not be made of the first occasions when such data can be linked for some pupils across the full five years of secondary schooling. As noted above, the DfEE Value Added pilot study suggested that it would not be until 2003 that the first of such analyses could realistically take place.

Second, this is the first time that such a comprehensive analysis of secondary schooling has been possible, the results outlined relate, therefore, only to one cohort of pupils. Further work on future years' data will be necessary to establish the robustness of the findings reported here.

Third, the 'excellence' of a school's performance is not synonymous with its *overall* examination performance. Aspects of particular interest, such as pupils' outcomes in individual subject areas such as English, Mathematics and Science have a particular importance; how well schools perform in these key areas has considerable relevance to pupils' later careers and should form a further focus for the evaluation of school performance.

Finally, schools do not exist simply as production lines for the generation of examination results; they are communities whose good health is enhanced by introducing pupils to a wide and liberal familiarity with the range and depth of our cultural heritage and its religious and philosophical underpinnings. This, together with the offer of opportunities to build on pupils' physical and artistic strengths, makes overall 'evaluation' of each school's particular strengths and weaknesses a 'multi-dimensional' act.

14. CONCLUSIONS

When all is said and done, however, much of the public debate about the comparative advantages of one form of educational provision as against another has been conducted on the basis of the very public issue of examination results. This paper has used some of the most recently available data, along with newer statistical methods, to shed further light on what is one of the most contested educational issues of our time. The conclusions are as follows:

- **The evidence reported here offers no support to the claim that ‘selective’ educational systems provide better GCSE examination results than comprehensive systems;**
- **The evidence indicates clearly that comprehensive systems of educational organisation are now delivering performance that is at least as good if not better than that achieved by selective systems.**
- **Selective systems of educational organisation, with the majority of their pupils in secondary modern schools, appear to perform less well overall than similar pupils in fully comprehensive systems.**
- Whilst, in the light of the cautions expressed in Section 13 above, the findings reported here are agnostic about the *size* of any advantage which comprehensive systems of organisation might have shown, it is unequivocal in its rejection of the alternative hypothesis that grammar schools (and their associated secondary modern schools) provide ‘better’ examination outcomes than those achieved by pupils in comprehensive schools and comprehensive systems of educational organisation . Almost all of the findings reported here point in the opposite direction

That this is a matter for continuing concern is illustrated by the DfEE’s Year 2000 Performance Tables for two ‘selective’ LEAs. These show, for example, that 13 of Kent’s 33 grammar schools are in OFSTED’s category E (denoting schools providing performance ‘well below’ that of comparable

schools) for GCSE (points score) performance, whilst a further eight are in category D. In Lincolnshire, another selective LEA, 7 out of its 15 grammar schools are in OFSTED's category E with a further two in category D.

15. RECOMMENDATIONS FOR ACTION

The issues discussed by this paper have been controversial for well over two decades, but debate has been somewhat muted until recently, given both the political climate and the lack of definitive studies designed to test alternatives to the claims advanced by supporters of selective education. The recent revival of interest is, at least in part, due to the possibility that parental ballots will see the end of selective education in some of those areas in which it still survives; but it is also a response to a new situation in which national data and appropriate methodologies are, for the first time, available to help inform this debate with stronger statistical evidence.

Under these circumstances it is regrettable that more information is not available from official sources to help inform the processes of parental choice. One very positive step would be to make Value Added data available for all secondary schools and to publish this in a helpful form alongside other material in the annual Secondary School Performance Tables.

The Department for Education and Employment (with OfSTED and QCA) already publish annually the very helpful Autumn Package which allows schools to make their own evaluations of their performance. What is needed is some more high profile publication of these evaluations along with guidance about what its findings show. Surely, this is the lead for which parents and others in the process of making decisions which can affect pupils' educational performance for years to come now have a right to expect from central government sources? To await some indeterminate point in the future when the conditions for doing so might be pronounced 'more perfect' is to put off the very

real advantages of using the available data *now* to do what the analyses presented in this paper have shown to be possible.

In a period when the merits of ‘selective’ systems of educational organisation are being debated more than at any time previously, it is important both to use the best sources of data available, and also the most relevant methodologies to carry out the appropriate comparisons between them. It is not that ‘absolute certainty’ can be achieved by acting in this way, but in the present state of affairs there is a very real need for what evidence there is to be made available publicly without recourse to ‘special pleading’ by advocates of one view or another. The time is now right for this. Only by so doing can we hope to bring greater clarity into debates about the merits of competing systems of educational organisation of secondary schools in England.

References

- AITKIN, M. AND LONGFORD, N. (1986), Statistical Modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, Series A, 144, 1.
- BOYSON, R. (1969), The essential conditions for the success of comprehensive schools, in Cox, C. and Dyson, A. (eds) *Fight for Education: a Black Paper*. London, Critical Quarterly Society.
- BRYK, A and RAUDENBUSH, S. (1992), *Hierarchical Linear Models*, Newbury Park, Sage.
- COX, C. and MARKS, J. (1980) , *Real Concern; An Appraisal of the National Children's Bureau Report on Progress in Secondary Schools*. London, Centre for Policy Studies.
- CROOK, D., POWER, S. and WHITTY, G. (1999), *The Grammar School Question: Review of research on comprehensive and selective education*, London, Institute of Education.
- DEPARTMENT OF EDUCATION AND SCIENCE (1983), *School standards and spending: statistical analysis*. Statistical Bulletin 16/83. London, DES.
- DEPARTMENT OF EDUCATION AND SCIENCE (1984), *School standards and spending: Statistical analysis a further appreciation*. Statistical Bulletin 13/84. London, DES.
- DEPARTMENT FOR EDUCATION AND EMPLOYMENT (1998), *Performance Tables for Secondary Schools*. London DfEE.
- DEPARTMENT FOR EDUCATION AND EMPLOYMENT (1998), *Value Added Pilot Study Supplement to Performance Tables for Secondary Schools*. London, DfEE.
- DEPARTMENT FOR EDUCATION AND EMPLOYMENT (1999), *Arrangements for Parental Ballots in respective of grammar schools*. London, DfEE.
- DEPARTMENT FOR EDUCATION AND EMPLOYMENT, QCA & OFSTED (1998, 1999), *The Autumn Package*, London, DfEE.

FITZ-GIBBON, C. (1997), *Value Added National Project; Final Report: Feasibility studies for a national system of Value Added indicators*. London, SCAA, COM/97/844

GOLDSTEIN, H. (1987), *Multilevel Modelling in educational and social research*, London, Griffin.

GOLDSTEIN, H. (1999), quoted in *Times Educational Supplement*, 26 November.

GRAY, J., JESSON, D. and JONES, B. (1984), Predicting differences in examination results between local education authorities: does school organisation matter, *Oxford Review of Education*, 10, 1, 45-68.

GRAY, J., JESSON, D. and JONES, B. (1986), The search for a fairer way of comparing schools' examination results, *Research Papers in Education*, 1, 2, 91-122.

GRAY, J., HANNON, V., JESSON, D., JONES, B. and RANSON, S. (1986), The Contexts Project: the use and interpretations of examination results as measures of school performance. *Final Report to the Economic and Social Research Council*, (pp. 52)

GRAY, J. and JESSON, D. (1987), Examination results and local authority league tables, in A. Harrison and J. Gretton (eds.) *Education and Training UK 1987*, Newbury: Policy Journals, (pp. 33-41).

GRAY, J. and JESSON, D. (1989), The impact of comprehensive reforms, in R. Lowe (ed.) *The Changing Secondary School*, Lewes: Falmer Press, (pp. 72-98).

GRIGGS, C. (1989), The New Right and English secondary education, in R. Lowe (ed) *The Changing Secondary School*. Lewes, The Falmer Press.

JESSON, D (1999), *Value Added Estimates of GCSE/GNVQ Performance in 1998 for Schools in the Technology Colleges Network*. London, Technology Colleges Trust, 15pp.

JESSON, D. (1997), *Value Added Measures of School GCSE Performance: Final Report*, London: HMSO (ISBN 0-85-522617-X, 44 pp.

LORD, R. (1984), *Value for Money in Education*. London, Public Money.

MARKS, J., COX. C. and POMIAN-SZREDNICKI, M (1983), *Standards in English Schools*. London, National Council for Educational Standards.

MARKS, J and POMIAN-SZREDNICKI, M (1985), *Standards in English Schools, Second Report*. London, National Council for Educational Standards.

MARKS, J., COX,C. and POMIAN-SZREDNICKI, M (1986), *Examination Performance of Secondary Schools in the Inner London Education Authority*. London, National Council for Educational Standards.

MARKS, J. (1991), *Standards in Schools: Assessment, Accountability and the Purposes of Education*. London, Social Market Foundation.

MORTIMORE, P. and MORTIMORE, J. (1986), Education and Social Class, in R. Rogers (ed) *Education and Social Class*, London, The Falmer Press.

OFSTED (1999), Introduction to *The New Handbook for the Inspection of Schools*. London, OFSTED

SCAA (1994), *Value Added Performance Indicators for Schools*, The School Curriculum and Assessment Authority, London: SCAA Publications (Ref: COM/94/151), ISBN 1 85838 045 6, 85 pages.