Evidence on the effects of selective educational systems

A report for the Sutton Trust

Robert Coe, Karen Jones, Jeff Searle, Dimitra Kokotsaki, Azlina Mohd Kosnin, Paul Skinner

CEM Centre, Durham University, UK
www.cemcentre.org

October 2008

Executive summary

The context of selection

1. There are 164 secondary maintained selective (grammar) schools in England, located in 36 Local Authorities (LAs). They arise from a complex history, and exist in the context of a wide variety of different kinds of secondary school (Chapter 1).
2. A number of arguments have been made for selection. These include the claims that it is appropriate for different types of pupil to have different kinds of education; that teaching can best be targeted at a narrow ability range; that grammar schools are meritocratic and socially redistributive by providing advantage for the bright but poor; that they are socially inclusive, as they keep the middle classes in state education; that the academic elite should be a priority for education; that grammar schools provide a beacon of excellence; that they achieve better academic results; and that selection operates elsewhere within the educational system (Section 2.1, p15).
3. Arguments against selection include the claims that selection tests are never fair or adequate; that ability is multi-dimensional and fluid; that the impact of failure on pupils not selected is unacceptable; that selection has an adverse effect on the primary school curriculum; that it is socially divisive; that selection compounds disadvantage; that it is the socially disadvantaged who should be a priority for education; that selection limits parental choice; and that selective systems produce worse academic results (Section 2.2, p20).

Existing studies

4. A number of studies have previously tried to compare the performance of pupils in selective and non-selective schools. Two major contributions in the 1980s (Steedman, 1980, 1983 and Marks et al., 1983, 1985) were followed by more recent interest (Jesson, 2000, 2001; Prais, 2001, Yang and Woodhouse, 2001). The advent of national pupil-level datasets allowed Schagen and Schagen (2003, 2005) and Atkinson et al. (2004) to advance our knowledge appreciably. Recent updates using data from the 1958 birth cohort National Child Development Study (NCDS) (Sullivan and Heath, 2002; Galindo-Rueda and Vignoles, 2004; Manning and Pischke, 2004) have also contributed. Some studies of Northern Ireland, the UK as a whole, and Australia are also relevant (Chapter 3).
5. Most of these studies suffer from limitations of methodology, data or interpretation; some are quite serious. In particular, their inability to control for other differences; problems with the quality of baseline or outcome data; issues in the calculation of value-added; inappropriate choice of the unit of analysis; failure to acknowledge the heterogeneity of selective systems; focus on cohorts that were educated in the 1970s; and researchers' apparent preconceptions all undermine the trustworthiness of their results (Chapter 4, p107).

6. Most of the existing studies report somewhat mixed results, with no clear advantage to either selective or non-selective systems as a whole. However, the majority of studies (and all of those we judge to be methodologically strongest) report that pupils who attend grammar schools do better than equally able pupils in comprehensives. This is true both for those that used national datasets and those based on the NCDS data. Some studies identify particular subgroups as benefiting most from attending a grammar school (Section 5.4, p131).

Our own analysis of national datasets

7. Just under 4% of 11-16 year-olds attend grammar schools. As well as being more able, they are also significantly less likely to be eligible for Free School Meals (FSM) than those in non-selective schools (Section 6.1.1, p137). This difference does not seem to be fully explained by their higher ability or their tendency to live in more socially advantaged areas (Section 7.2.9, p183).
8. In terms of school-level characteristics, grammar schools are very different from other schools. All have sixth forms, compared with about half of non-grammar schools. Fewer than 10% of non-grammars are single-sex schools, compared to three-quarters of grammars. Grammar schools contain higher proportions of Specialist schools and Foundation schools (Section 6.1.2, p138).
9. Pupils who attend grammar schools do not all live in the LA of the school they attend. Nationally, about 20% of grammar school pupils come from outside the LA; for some LAs, this figure is as high as 75% (Section 6.3.1, p142). Some 80 LAs have more than 1% of the pupils who live in their area attending grammar schools, compared with only 36 LAs that actually have grammar schools of their own (Section 6.2.2, p139). Across England as a whole, one third of the wards in the country (33%) house at least one pupil who attends a grammar school (Section 6.2.1, p139). The concept of a 'selective' LA, whose performance can be isolated, is therefore rather problematic.
10. Different qualifications taken at KS4 are not of equal difficulty, and the points awarded to 'equivalent' qualifications do not necessarily reflect this. In some subjects, students are systematically getting better grades than those same students do across their other subjects. Overall, more able pupils tend to take harder qualifications, and those in grammar schools even harder still. Any comparison of grades achieved should therefore take account of these differences (Section 6.4, p144).
11. We have developed a way of defining the 'creaming' effect of any given grammar school on each non-grammar school (Section 7.1, p153). A relatively small number of schools are substantially creamed: 161 schools (5% of non-selective schools nationally) lose more than 20% of their potential pupils to grammar schools. Three-quarters of these schools are in just four LAs. Just under one-third of the non-selective schools in the country (32%) lose between 0 and 1% of the pupils they might have had, with a further third (35%) losing between 1% and 20%. Throughout the country as a whole only about one-quarter of non-selective schools (28%) lose no pupils at all to grammar schools (p157). This far-reaching but low-level impact of selection is very different from the traditional picture of

self-contained 'selective' and 'comprehensive' areas, with grammar and secondary modern schools on the one hand, and comprehensive schools on the other.

12. We have also developed a method for calculating 'selectivity': the extent to which schools discriminate academically and socially between the pupils they take and those, living in the same neighbourhoods, whom they do not (Section 7.2, p161). Not surprisingly, grammar schools are substantially more academically selective than other schools, though, surprisingly, there is actually some overlap. Grammar schools are also more socially selective than other schools, but here the overlap is much bigger; the most socially selective state schools in the country are 'non-selective' schools. These socially selective schools are more likely to be Voluntary Aided or CTCs, to be single sex, faith schools, larger than average and drawing from more competed wards (p178).
13. In comparing the performance of pupils in selective and non-selective schools, a number of choices have to be made. These include how the different Key Stage 4 outcomes should be treated; what kinds of factors should be taken into account in order to make comparisons fair; what kinds of statistical models should be used; and which groups should be compared. Implications of different choices are considered (Section 8.1, p187).
14. In terms of raw KS4 (GCSE) results, it is clear that pupils in grammar schools do much better. This advantage remains, although the difference is smaller, if consideration is limited to pupils who achieved level 5 or higher in each of mathematics, English and science at KS2 (Section 8.2, p195).
15. Regression and multilevel analyses were conducted on the national pupil data (section 8.3, p209). Propensity Score Matching was also applied (section 8.5, p222). Most of these analyses suggest that pupils in grammar schools do a little better than similar pupils in other schools, with the difference somewhere between zero and three-quarters of a GCSE grade per subject. In general, the more factors introduced into the model, the smaller the difference. In particular, the inclusion of school composition variables reduces the grammar school advantage considerably (p213). The choice of different statistical models also makes a difference to the outcome, as does the use of different outcome measures. On the other hand, the choice of different comparison groups does not seem to make much difference to the results. The schools that are affected by grammar schools, in terms of losing pupils to them, are performing no differently from all other schools (Section 8.3.4, 215). Although these analyses indicate that grammar school pupils appear to make greater progress from KS2 to KS4 than other pupils, we also find that these same pupils were already making more progress from KS1 to KS2 (ie in their primary school). This suggests that there may be important but unmeasured differences between grammar and non-grammar school pupils and somewhat undermines our confidence in these estimates of a 'grammar school effect' (section 8.4, p220).
16. Overall, therefore, we find that although many of our analyses identify a small positive advantage in GCSE achievement for pupils at grammar schools, there are good reasons to be cautious of describing this as a grammar school 'effect'. At least a part of this difference is likely to be a

result of inadequate data and bias in the evaluation designs available to us.

Contents

List of Tables

List of Figures

PART I
INTRODUCTION

1. Introduction: The context of selection

The 164 grammar schools in England have been the subject of enormous political controversy and research attention. To some they are a beacon of quality in a sea of mediocrity; to others they are an iniquitous reinforcement of social privilege and segregation. Views are strongly held on both sides and at times when their existence has been under question, campaigning has been impassioned. Although at the time of writing in England the existence of grammar schools seems not to be high on the political agenda, given its turbulent history there can be little doubt that it will one day return to prominence. For anyone with an interest in educational policy, questions about the role and impact of selective schools are perennially relevant.

Questions concerning selection for grammar schools and entry to comprehensive schools appear to be of equal interest to a researcher's point of view. Few areas of education policy in this country have been the subject of more research. Despite – or perhaps because of – this, there is little agreement about the impact of grammar schools on the educational achievements of the pupils who attend them, or on those who do not. In the academic arena, as in the policy arena, opposing views are strongly debated.

In this chapter we present a brief survey of the context of selection, describing the situation as it is now and how it came to be so. We try to clarify the status of different types of school within England and of different Local Education Authorities (LEAs). We then provide an overview of what follows in the rest of the report.

1.1. The context of selection in England

1.1.1. A brief history of English schools

Some 1000 years ago the first English schools were set up to teach Latin grammar. Eventually some of these evolved to become state funded grammar schools whilst others became privately funded independent schools. The oldest existing state grammar school, Reading School, was founded in 1125, becoming Reading Grammar School in 1486. In the period between about 1500 and 1850 a number of Charitable Foundations set up the endowed grammar schools, including those in the King Edward VI Foundation in Birmingham.

In 1856 a government department of education was established, as the idea that schooling should be available to all, not just to children of parents who could buy private education, began to gain political currency. At the same time the government began to provide finance to schools. The Education Act of 1870 (The Forster Act) can be regarded as the beginning of the modern system of education in England. In that Act the government set up elementary schools, run by school boards, to supplement schools run privately by churches and others. Attendance to age 12 became compulsory for all children, but fees were still charged. Elementary education effectively

became free of charge with the 1891 Education Act. The debate about secondary education and its availability to all dates from the same period, though the term 'primary' was yet to be used.

In 1902 the Balfour Act established the local education authorities (LEAs). By this time there was a grammar school in most major centres of population either *maintained* by the state or *endowed*, funded through charitable foundations.

The beginning of the 20th Century marked the beginning of government involvement in education as a policy making body that could direct the LEAs through legislation. There was a perception of a three class education system comprising

1. the public schools (Eton, Harrow etc),
2. the grammar schools (for the academically able) and private schools (for children of parents who could buy a secondary education), and
3. the elementary schools (for children of the working class). Attendance at the elementary schools was very sporadic.

By 1917 a formal examined qualification, the School Certificate, was linked to grammar schools. Grammar schools were now seen as the route to higher education and professional careers, particularly for children from working class backgrounds who won a grammar school scholarship or free place. Growing unease over the disparity between elementary schools and grammar schools, however led to the Hadow Report of 1926, which recommended transfer at age 11 and introduced the idea of primary and secondary education. In the years leading up to the Hadow Report the terms secondary education and grammar schools had been synonymous. Hadow proposed a new type of secondary school to be called modern schools, for adolescents to attend between the ages of 11 and 15 if they did not attend a grammar school. However, it was not until the 1944 Education Act that this recommendation was acted on. The Spens Report of 1938 recommended there be a third type of school, a technical high school, and that there should be parity of status between the three types of school.

The two decades before World War II saw a developing unease with the perceived meritocratic or elitist system of education. However, the War itself and hopes for the post war society, were stimuli for large scale reconstruction of the education system. The white paper of 1943, *Educational Reconstruction*, noted that

> The keynote of the new system will be that the child is the centre of education and that, as far as is humanly possible, all children should receive the type of education for which they are best adapted… academic training characterised by the existing secondary schools (i.e. grammar schools) would be unsuitable for the majority of pupils. (Quoted from McCulloch, 2002).

1943 also saw the publication of the Norwood Report, a report to The Board of Education on curriculum and examinations, which supported the tripartite system of grammar schools, secondary modern schools and technical schools and this was enshrined in law in 1944 in the Butler Act. In practice few technical schools were ever set up, due to the difficulties of identifying technical aptitude in a 10 year old child. Entry to the grammar schools was through the 11+ test, which would be passed by about 25% of children, the

remainder attending the vocationally orientated modern schools. As Crook *et al* (1999) observe,

> The orthodoxy that intelligence was measurable by psychometric tests offering a neutral means of assessing the aptitudes of children from deprived backgrounds and allocating them to appropriate schools, had dominated a generation of educational thinking.

The Butler Act also marks the beginning of the alignment of right versus left (Conservative vs. Labour) in the selection vs. comprehensive debate. Butler was a conservative, but it was the post war Labour Government that implemented the act. The beginnings of a comprehensive movement can be traced back to the 1920s but in the 60 years subsequent to the 1944 Act, there has never been a clear, black and white, dichotomy between the left and right in politics and comprehensive versus selective education. Again, to quote from Crook *et al* (1999),

> Grammar schools, a number of which enjoyed reputations going back to the sixteenth century, had been successful in producing a formidable number of Labour politicians at both the national and local levels.

During the 1950s and 1960s the tripartite system remained in place in most LEAs. There was some experimentation with comprehensive schooling but in 1965 there were over 1000 grammar schools in England and Wales.

In 1965 the Labour Government, under Education Secretary Tony Crosland, issued circular 10/65, which marked the beginning of the debate, which is still continuing. Circular 10/65 requested (i.e. did not dictate *must*) that all the LEAs draw up local plans for schooling post age 11 along comprehensive lines. There was no single model and the Department of Education and Science accepted a number of plans that sought to soften selection rather than remove it altogether thus suggesting to some that the coexistence of grammar schools and comprehensive schools was possible. (Crook *et al*, 1999)

In 1969 the Labour Government tried to strengthen the legislation in a similar manner to Scotland, where a fully comprehensive system had been put in place. The 1969 general election intervened, and a conservative government was returned, in which Margaret Thatcher held the Education portfolio. Many find it ironic that she approved more plans for comprehensive reorganisation than any Labour minister. However, she withdrew Circular 10/65 and replaced it with circular 10/70 which permitted the coexistence of grammar schools, modern schools and comprehensives schools. The extent to which various LEAs went fully comprehensive has been much debated, with many seeing some comprehensive schools as simply being modern schools renamed; it was notable that most of the comprehensive schools with sixth forms were formed from former grammar schools.

In 1974 a Labour Government was returned to power; circular 10/70 was withdrawn and replaced with 4/74, once again requesting LEAs to submit plans for comprehensive reorganisation. At this time 26% of 10 year olds still sat the 11+ examination. Many Conservative Authorities chose to defy the legislation and further legislation was introduced, but not enacted as James Callaghan as Prime Minister and Shirley Williams as Secretary of State for Education had begun *The Great Debate*. (Ruskin College, 1976). A DES report in 1978 confirmed the Labour Government's unwillingness to differentiate between the genuinely comprehensive and partially comprehensive solutions adopted by a number of LEAs.

Margaret Thatcher's Conservative Government came to power in 1979 and introduced an agenda of *Choice and Diversity* into education. Part of this agenda was the introduction in 1981 of the Assisted Places Scheme, in which academically able children could gain places in independent schools with their fees paid by the state.

By 1980, 80% of children age 11-16 were in schools called comprehensive schools, with 5% in state grammar schools and 6% in the independent sector (Gray & Jesson, 1989) but these authors noted it took 5 years for a school to become comprehensive in the 11-16 year groups after reorganisation. Gray & Jesson (1989) point out that around 1980 over 1 in 10 children were being educated in schools that were predominantly selective, and there was widespread understanding, approval and support for what such schools had stood for and still stand for. Gray & Jesson concluded that although comprehensive reforms can claim to have delayed some aspects of selection to 14, 15 or 16 years of age, they have not abolished them. In summarising the studies to date (1989) that had compared comprehensive and selective systems, they concluded,

> … if the major purpose of comprehensive reorganisation was to secure higher performance levels among more able pupils than the previous selective system probably achieved, then it must be said to have failed. If the primary purpose was to ensure that all pupils had the opportunity to achieve the levels of qualifications of which they were apparently capable then the evidence suggests some modest success.

Gray & Jesson's tentative remarks here set the scene for the later, bigger studies that were to follow.

Legislation in 1988 introduced two new types of self-governing secondary schools; the City Technology Colleges and the Grant Maintained Schools, state funded but independent of LEA control. The legislation allowed schools to *opt out* of LEA control, and introduce their own admissions policies thus partially reintroducing selection. *Choice and Diversity* was formalised in a White Paper in 1992, which introduced another new type of school, the Specialist School, with specialisms in particular areas of the curriculum, such as technology, music and languages. These schools could select up to 10% of their intake by the children's aptitude for the specialism.

The term *creaming* came into use in about 1980 as a way of describing how certain schools attracted the more able children, or their parents, thus reducing the likelihood of other schools having such children amongst their pupils:

> What is clear is that some comprehensives have been subjected to two forms of creaming of their most able potential clients: by selective schools remaining in the public sector and by selective schools in the private sector. Some of these latter schools had previously been part of the direct grant system which had provided 11+ places for scholars alongside grammar schools. Given the choice of becoming part of the comprehensive system or going independent many had done the latter taking their pupils with them. (Gray & Jesson, 1989, p74)

Some writers (e.g. Griggs, 1989) have suggested that the Thatcher government and in particular her education ministers, Keith Joseph and Rhodes Boyson, were very much influenced by right wing *think tanks* and

right wing writers in general. In particular, Griggs cites the Institute of Economic Affairs (from 1957) and the *Black Papers* (from 1969) as influencing the Centre for Policy Studies, and groups such as the Parental Alliance for Choice in Education and the Campaign for Real Education whose various pressure activities led up to the 1988 Act and the creation of City Technology Colleges and the Grant Maintained Schools. Boyson claimed that standards were falling in education and that bright working class children were missing out by attending a neighbourhood comprehensive rather than a selective grammar school.

In 1996, Prime Minister John Major offered a grammar school in every major city if that is what the people wanted. However, when his government brought in new legislation the emphasis was on increasing the number of specialist schools with a model for secondary education comprising 50% Grant Maintained; 30% specialist and 20% LEA comprehensives. Gillian Shepherd (Secretary of State in Major's government) did not get the chance to enact her Bill as Tony Blair's New Labour Party won the 1997 general election.

David Blunkett, when shadow Secretary of State, is famously on record as stating 'no selection under a Labour government.' (Labour Party conference, 1995.) This was later corrected to claim that he *meant* to say 'no *more* selection.' Tony Blair meanwhile was saying he would support parents' views, and introduced the legislation that would allow the remaining grammar schools to become comprehensive if that is what the parents wanted. Only one ballot under this legislation has ever taken place, and the parents of Ripon (Yorkshire) voted to keep their grammar school. As Blair said in 1997,

> Let me say this about school structures; I have no intention of waging war on any schools except failing schools. As far as the existing 160 *[sic]* grammar schools are concerned as long as parents want them, they will stay. We will tackle what isn't working not what is. (Speech at Barber Institute, University of Birmingham, quoted in Crook *et al*)

He later (1999) pointed out that the Conservative Party had closed more grammar schools than any Labour Government.

In 1998 the School Standards and Framework Act was introduced indicating that New Labour supported the growth of the specialist schools and saw selection by aptitude as the way forward for secondary education. Whilst New Labour pursued its policy of encouraging all secondary schools to apply for specialist status, in the Education Act of 2002 it introduced a new type of school, to be known as Academies, aimed particularly at replacing inner city schools that were perceived to be failing their pupils. By the end of 2005, there were 27 Academies in England, but the debate about the most appropriate type school to give life chances to children from poor backgrounds continued. In his bid to become leader of the Conservative Party in 2005, David Davis said he would create 20 new grammar schools in the cities, saying they would revolutionise education in the inner cities through creating equality of opportunity by giving poor children the same choice that well off children have always had. Mr Davis said the new grammar schools would be selected from those in the current city academy programme. Mr Davis was unsuccessful in his bid.

It is appropriate at this point to review the types of secondary school that exist today in the state sector in England and their admission policies.

1.1.2. Different schools and their admissions policies

Selection and admissions policies to secondary schools are closely related. It will be helpful here to clarify the types of secondary school that exist in England in 2008 and their admissions policies and who has responsibility for admissions.

Within each Local Authority (LA) in England there is a Directorate of Children's Services which has responsibility for schools. Much of the funding from the DCSF (formerly DfES) for schools now goes directly to schools rather than the LA, so their role is more strategic than operational.

There are six governance types of maintained secondary school (maintained means the operational costs of running the school are met by the state). The first two of these, Academies and City Technology Colleges, are essentially independent of LA control.

Academies – are all ability schools established by sponsors from business, faith or voluntary groups working in partnerships with central Government and local education partners. Private sponsors and the DCFS provide the capital costs, whilst running costs are funded by the DCFS.

Academies are located in areas of disadvantage where they have replaced weak or failing schools. Academies have flexibility in their curriculum to meet local needs. Sponsors aim to make a break with a culture of low aspiration which afflicts many inner city schools and to counteract the impact of deprivation on education in the community. All Academies have Specialist School status. By the end of 2007 there were 87 Academies, with a target of 200 by 2010.

City Technology Colleges – were established as part of the Education Reform Act 1988; fourteen colleges (CTCs) and one city college for the technology of arts (CCTA) were opened in urban areas across England in the period 1988 to 1993. The purpose of the CTCs was to provide a broadly based secondary education with a strong technological element offering a wider choice of secondary school to inner city children aged 11-18. They were set up using private finance but with operational costs met directly from the government. The governing body is responsible for admissions. Although they teach the National Curriculum, these schools focus on technology, science and mathematics and offer vocational as well as academic courses to post 16 students. Some CTCs have changed status to Academies and in 2006 only 11 CTCs remained.

Community Schools – (previously known as County Schools). The LA is responsible for employing the staff of the school, for the school buildings and land, and for admissions to the school.

Foundation Schools – (mostly previously known as Grant Maintained schools). The governing body of the school employs the staff and has a major responsibility for admissions to the school. The buildings and land are owned by the governing body or a charitable foundation.

In the Education Act 2006, the government introduced a new type of Foundation School, known as a Trust School. A Trust school forms a charitable trust with an external partner such as a business or educational charity, with the aim of raising standards and exploring new ways of working. Trust schools are able to manage their own assets, employ their own staff and set their own admission arrangements. In 2007 there were 69

Trust "pathfinder" schools coming from the primary, secondary and special schools sectors and there were plans to substantially increase this number.

Voluntary Aided – (many of these are church schools). The governing body of the school employs the staff and has responsibility for admissions to the school. The buildings and land are owned by a charitable foundation. The King Edward VI grammar schools are voluntary aided.

Voluntary Controlled – (most of these are church schools). The LA is responsible for employing the staff of the school and for admissions to the school. The school buildings and land are mostly owned by a charitable foundation.

Within this structure there are other possible types of school. These include Specialist Schools, Faith Schools and Selective (Grammar) Schools.

Specialist Schools – were introduced in 1994 and the first Specialist Schools were in technology. The incoming Labour government in 1997 endorsed and developed the concept and there is now a diverse range of ten specialist areas; the ten categories of specialist school are: Technology; Language; Arts; Sports; Business and Enterprise; Engineering; Mathematics & Computing; Science; Humanities; Music. The government has a policy of encouraging all secondary schools to seek specialist status and the initial target of 2000 was passed early in 2005. The number of Specialist Schools is now approaching 3000, or about 85% of all secondary schools involving over 2.5 million school children with a target now of all secondary schools becoming specialist in some area. This is part of the government agenda of creating greater flexibility and freedom for schools; for example Specialist Schools can raise funds through private sector sponsorship. The belief is that through focussing on their specialist area, Specialist Schools raise standards across the whole curriculum.

Faith Schools – some maintained schools are also called Faith Schools as most of their staff and pupils practise a particular religion, which is an integral part of the school ethos. Most faith schools are Christian in nature but others, such as Jewish and Muslim Schools, are becoming established. Faith Schools are generally voluntary controlled or voluntary aided. Faith schools generally admit pupils of religious affiliation to the faith, although many accept a proportion of others and in some schools this is not a necessary requirement.

Within the admissions policies of many of these schools there is an element of selection. A school is over subscribed if it has more applicants for places at the school than it has places available. Thus those responsible for entry to over-subscribed schools must have a way of deciding which children are admitted and which are not. At the time of the Education Reform Act, 1988, 15% of schools were their own admission authority; in 2004 this had doubled to 30% (West, Hind and Pennell, 2004). Over-subscribed schools are mostly foundation and voluntary aided schools and such schools are thus in a position to select pupils.

Within the 1998 Act, there was a Code of Practice through which the government aimed to make schools admissions policies both transparent and fair. This has recently been revised and the new Schools Admissions Code came into force in February 2007. All maintained schools are bound by law to follow the Code, however CTCs and Academies are not maintained schools, and their admission arrangements require approval by the Secretary of State

for Education. Every school is required to have a published admissions policy which includes criteria on how places will be allocated when a school is over-subscribed. The Code aims to promote equity, fair access to educational opportunity and increasing parental choice. Within the 1998 Act there was variation amongst schools and LAs but these criteria included siblings already at the school, distance to school, medical and social needs, catchment area, first preference, special educational needs and feeder schools (West *et al*, 2004). However, the new Code has tightened up on some of these with the overall aim of ensuring fair admissions systems. For example interviewing of applicants or their parents is prohibited, as is any attempt by parents to influence a decision by offering support of some kind to the school.

However, what is of particular interest in this study is that the 1998 Act permitted any school to select 10% of its pupils on aptitude if the governing body is satisfied the school has a specialism. This represented a development of the Specialist Schools concept, since they came into being following the 1992 White Paper *Choice and Diversity*. In 1999 there were about 400 Specialist Schools; this number exceeded 1300 in the summer of 2004. Subsequent to 2004, schools acquiring Specialist Status can not select on aptitude.

In the 1998 Act, ability is said to be either general ability or ability in any particular subject or subjects. The Act does not define aptitude but according to the Code of Practice, '… a pupil with aptitude is one who is identified as being able to benefit from teaching in a specific subject or one who demonstrates a particular capacity to success in that subject.' As noted by West *et al* (2004), '…it is not clear how demonstrating a capacity to succeed differs from ability.' Thus there was an argument that the government supported grammar school type selection, through this covert terminology of aptitude but this has now been dropped for the post 2004 Specialist Schools, a situation that some Conservative politicians have described as an extraordinary anomaly.

Following the Education Reform Act of 1988 and the Schools Standards and Framework Act 1998, there exists a quasi-market in secondary education in which schools have some autonomy over which children are admitted and parents can express a preference as to which school they wish their child to attend (West *et al*, 2004). The focus in this quasi-market is on which pupils are admitted rather than on which pupils are selected. The revised Schools Admission Code 2007, aims to make this a fair market but quasi-markets in education are not unique to England. They have been evolving in many countries across the western world for many years. For example, Whitty, Power and Halpin (1998) have discussed these developments in the USA, Australia, and New Zealand as well as in Sweden and England.

1.1.3. Grammar schools today

There are 164 state funded, selective grammar schools in England. The number has been static since 1999 when the two grammar schools in Bristol first accepted a comprehensive intake at year 7. These 164 schools are selective in the sense that they are designated as such by the DCSF. This means they have some sort of entrance test, which pupils must pass in order to enter year 7 of the school. There are, however, another 24 state funded schools in England that have the word *grammar* in their title. These are former grammar schools which are no longer designated as selective by the DCSF, but as comprehensive. Some of these schools, indeed comprehensive schools

in general, may have some sort of restrictive entrance policy as discussed in Section 1.1.2 (p8). Many independent schools (46) also have the term *grammar* in their titles. By virtue of being state funded, all of the 164 grammar schools come within the jurisdiction of a Local Authority, but the LA isn't necessarily responsible for the entrance test or the admissions policy of a particular school.

Of the 150 LAs in England, 36 have at least one grammar school whereas three LAs, Buckinghamshire, Lincolnshire and Kent, have more than ten. It is thus necessary to consider the extent of selection within an LA; should an LA be described as fully selective, partially selective or, if it has only one grammar school, as marginally selective?

Much of the debate on grammar schools and comprehensive schools has used the concept of a selective area. We show later in this report that such an area is difficult to define due to the great variation in the percentage of selection within an LA as well as where the grammar schools are located across the country. Some of the debate has referred to selective systems rather than selective areas. The term selective system may recognise the problem in pinning down a selective area geographically but it is still difficult to define. We shall discuss this further as a result of our own analysis in Chapter 6, but we shall say here, a selective system is an ill defined geographical area in which at least one grammar school is located, to which children from the area may apply for schooling post age 11.

Table 1 (p12) shows the number of secondary modern and comprehensive schools in the 36 LAs that have at least one of the 164 English state selective schools. There are no selective schools in Wales or Scotland. The number of independent schools in each of the LA areas is also shown. The final column shows percentage of students in each LA attending grammar schools as a measure of the level of selection. They are calculated from the number of pupils on the school roll in the DfES National Pupil Database (2006) (only mainstream schools have been included).

Table 1 indicates the wide variation in the level of selection between LAs but it also notable that only Buckinghamshire and Trafford are *totally selective*, in that they have no comprehensive schools. Some LAs are *partially selective* in that they maintain a bipartite system of grammar schools and secondary modern schools but these exist together with comprehensive schools. There are 15 LAs with at least one secondary modern school and one comprehensive school but the number varies considerably between LAs across the country. There are 19 LAs with at least one grammar school and no schools designated as *secondary modern*.

Table 1 also illustrates the wide variation in geographical location of the grammar schools in England. Thus some politicians, particularly some Conservatives continue to argue for more grammar schools. However, Conservative policy on selection became somewhat confused during the Spring of 2007. The official line from leader David Cameron and his shadow education secretary David Willetts is that a widespread extension of academic selection is ruled out and there will be no more grammar schools. They support the governments Academy programme but want the 10% selection by aptitude restored to all Specialist Schools. If that were to happen and the target of all secondary schools becoming Specialist Schools were met, than all secondary schools would have some control over who they chose to admit.

Table 1: LAs containing grammar schools

LEA	Total schools	Independent schools			State schools				
		Selective	Non-selective	Total independent schools	Grammar	Secondary modern	Comprehensive	Total state schools	Percentage selection
Slough	11	0	0	0	4	6	1	11	44
Buckinghamshire	47	10	3	13	13	21	0	34	43
Trafford	23	4	1	5	7	11	0	18	42
Kent	156	32	23	55	33	47	21	101	32
Sutton	16	2	0	2	5	0	9	14	32
Southend-on-Sea	19	2	2	4	4	4	7	15	31
Medway	24	2	2	4	6	10	4	20	31
Torbay	11	0	3	3	3	0	5	8	30
Bexley	17	0	0	0	4	5	8	17	27
Wirral	26	2	2	4	6	5	11	22	27
Lincolnshire	76	5	6	11	15	31	19	65	26
Reading	15	5	2	7	2	0	6	8	26
Poole	11	2	0	2	2	3	4	9	25
Bournemouth	12	1	1	2	2	6	2	10	22
Kingston upon Thames	15	3	1	4	2	8	1	11	20
Plymouth	19	1	1	2	3	0	14	17	15
Gloucestershire	61	12	5	17	7	0	37	44	14
Calderdale	20	3	2	5	2	0	13	15	13
Barnet	37	11	6	17	3	0	17	20	12
Birmingham	96	13	7	20	8	0	68	76	10
Warwickshire	45	7	1	8	5	11	21	37	10
Telford and Wrekin	19	2	2	4	2	1	12	15	9
Bromley	26	3	4	7	2	0	17	19	8
Redbridge	21	3	0	3	2	0	16	18	8
Lancashire	121	16	20	36	4	2	79	85	7
Enfield	20	3	0	3	1	0	16	17	6
Walsall	25	1	4	5	2	0	18	20	6
Stoke-on-Trent	18	1	0	1	1	0	16	17	6
Wiltshire	43	9	5	14	2	4	23	29	6
North Yorkshire	56	9	4	13	3	3	37	43	6
Kirklees	36	4	7	11	1	0	24	25	5
Wolverhampton	22	3	0	3	1	0	18	19	4
Essex	100	9	10	19	4	0	77	81	4
Liverpool	40	4	4	8	1	0	31	32	3
Cumbria	60	7	10	17	1	0	42	43	2
Devon	65	14	14	28	1	0	36	37	2
Totals	**1429**	**205**	**152**	**357**	**164**	**178**	**730**	**1072**	

Early in 2007, the right wing *think tank* the Centre for Policy Studies, carried out a survey of over 1000 people and reported that over 75% of them believed bright children would do better if taught separately and that a selective system or streaming would also help academically weaker children. The author of the report, Lord Blackwell, said grammar schools used to provide a route out for poor children but what we have now is selection by postcode rather than selection by ability. He claimed all children, particularly poor children, would benefit from selection saying most comprehensive schools cannot provide a large enough peer group to provide a challenge and peer group pressure for the most able (the top 5%). In his report, one of Lord

Blackwell's proposals is that state schools be enabled to opt to be fully selective. However, in response, Sir Cyril Taylor, of the Specialist Schools and Academies Trust, said he was totally against bringing back selective grammar schools in all areas highlighting that we need good schools for everybody. Sir Cyril described the existing 164 grammar schools as having effectively become "free independent schools" because they were not serving brighter children from poorer backgrounds. In May 2007, David Willetts made a major speech on Conservative Party policy on secondary education, supporting the Academies programme and rejecting the creation of any more grammar schools, although the existing 164 would remain. Willets said "the 11 plus entrenches advantage" and that "we must break free from the belief that academic selection is any longer the way to transform the life chances of bright poor kids". David Cameron reacted to dissent on his back benches to this speech by saying those wanting to build more grammar schools are delusional. Although Lord Blackwell's survey had found that about 40%% of parents would send their child to a selective school, Willetts described it as "fantasy" to say selection at age 11 could be fair. Simultaneously however, some back bench Labour politicians were hoping the government would abandon the Academies programme, seeing it as "privatisation" of state education.

Thus while the debate about selection continues, the 164 existing grammar schools remain in place, unlikely to grow in number, and without any uniformity in their distribution across the country. So given that these grammar schools do select their pupils at age 11, we raise the question as to what extent it makes any sense to talk about a selective system. This raises the question as to whether the LA is a meaningful unit to measure the level of selection, as the number and types of schools vary considerably. It also raises the difficulty of comparing the performance of children in selective areas of the country with children in non-selective areas, as these areas are clearly difficult to define. As there is no consistency between areas that have at least a degree of selection, it suggests that comparisons can only be made between one area of the country and another; the scope to make any generalisations is at least limited.

Much of the debate on grammar schools and comprehensive schools has used the concept of a selective area. We show later in this report that such an area is difficult to define due to the great variation in the percentage of selection within an LA as well as where the grammar schools are located across the country. Some of the debate has referred to selective systems rather than selective areas. The term selective system may recognise the problem in pinning down a selective area geographically but it is still difficult to define. We shall discuss this further as a result of our own analysis in Chapter 6, but we shall say here that a selective system is an ill defined geographical area in which at least one grammar school is located, to which children from the area may apply for schooling post age 11.

1.2. Overview of Report

This report is broken into four parts. In the second chapter of Part I we present a summary of the arguments that have been made for and against selection.

Part II of the report contains three chapters that present a literature review and discussion. Chapter 3 presents the review of major empirical studies that

have been carried out to compare pupil attainment in selective and comprehensive school systems. This is followed in Chapter 4 by a discussion of the limitations and problems we are faced by when attempting to evaluate selective systems. With these in mind we then proceed in Chapter 5 to a discussion of the evidence presented in the literature review.

In Part III we present our analysis of the national pupil datasets in three chapters. Chapter 6 looks at the differences between selective and non-selective schools. This chapter includes a more detailed analysis of issues introduced in 1.1.3 surrounding the definition of selective systems and ultimately questions the use of the LA as the unit for analysing selective systems.

Chapter 7 looks at the relationships between selective and non-selective schools in terms of the 'creaming' effect of the former on the latter and the ways in which they are 'selective' in their intakes. We attempt to answer the question, 'is there some other way we can define selective units of schools or areas, if not the LA?' by considering the extent to which non-selective schools are 'creamed' by selective schools. In this section we also attempt to look at patterns of selection and segregation in both grammar and non-grammar schools.

In Chapter 8 we continue the data analysis with a comparison of the performance at Key Stage 4 of selective and non-selective schools.

Finally, in Part IV, we present an overview of the report in Chapter 9 in the form of a summary and discussion of the analysis of the National Pupil Dataset. We also compare our results with previous studies and discuss implications for policy and for further study.

2. *Arguments for and against selection*

In this chapter we present a summary of the arguments that have been made for and against selection, for reference later when considering evidence that supports or counters them. The arguments include ideas about psychology, social and economic reasons, as well as ideas as to the best composition of classes and schools that is the basis of the whole topic being considered. These arguments stand or fall by the evidence of the outcomes, which is, as we have said, difficult to interpret and inconclusive at present.

For this chapter alone we ignore the difficulties in defining a "selective system" that was described in the previous chapter and use the term in a general sense without reference to actual schools or areas of the country.

2.1. *The case for selection*

We consider first the arguments for selection, beginning with the ideas behind the tri-partite selective system introduced after World War II, about social mixing and inclusion together with an economic argument. This is followed by arguments as to whether the system actually works, opinions about good composition of classes and schools and whether there is evidence from the outcomes in support.

2.1.1. Appropriate (different) education for different types of pupil

The need for selection enshrined in the 1944 Butler Education Act was derived from the ideas in The Norwood Report (1943) from which we quote in this section. The original concept enshrined in the Act was that pupils of different abilities required different types of education. Essentially the belief was that an academic curriculum, based on abstract reasoning, was not appropriate for all:

> In the secondary stage the attempt is made to provide for such special interests and aptitudes the kind of education most suited to them; they may have begun to indicate themselves at least roughly in the last phases of primary education, or they may not declare themselves in such degree as to deserve attention till a different kind of education is encountered. It is the business of secondary education first to provide opportunity for a special cast of mind to manifest itself, if not already manifest in the primary stage, and secondly, to develop special interests and aptitudes to the full by means of a curriculum and a life that is best calculated to this end.

Norwood went on to argue there are roughly three types of pupil, hence the need for a tripartite system. Those who could benefit from a grammar school education were described as:

> The pupil who is interested in learning for its own sake, who can grasp an argument or follow a piece of connected reasoning, who is interested in causes whether on the level of human volition or in the material world, who cares to know how things came to be as well

> as how they are, who is sensitive to language as expression of thought, to a proof as a precise demonstration, to a series of experiments justifying a principle; he *[sic]* is interested in the relatedness of related things, in development, in structure, in a coherent body of knowledge.

Pupils who would benefit from a technical education were described as:

> Having interests and abilities that lie markedly in the field of applied science or applied art. The boy *[sic]* in this group has a strong interest in this direction and often the necessary qualities of mind to make it his life's work at whatever level of achievement. … To justify itself to his mind, knowledge must be capable of immediate application and the knowledge and its application that appeal to him are concerned with the control of material things.

Those who would benefit from a secondary modern school education were described as follows:

> The pupil deals more easily with concrete things than with ideas. He *[sic]* may have much ability but it is in the realm of facts. He is interested in things as they are; he finds little attraction in the past or in the slow disentanglement of causes or movements. His mind must turn its knowledge or its curiosity to immediate test, and his test is essentially practical….because he is interested only in the moment he may be incapable of a long series of connected steps; relevance to present concerns is the only way of awakening interest, abstractions mean little to him.

In summarising his proposal for reorganisation of the whole secondary education system Norwood said:

> Under such a reorganisation all children would have the opportunity of the education best suited to them; for variety of type and alternative courses within the type are essential to any satisfactory system of secondary education. … To the three types of school parity of conditions should be accorded; parity of esteem must be won by the schools themselves. Such a reorganisation offers equivalence of opportunity to all children in the sense to which it has valid meaning; namely the opportunity to receive the education for which each pupil is best suited for such time and to such a point as is fully profitable to him.

Ignoring the apparently casual sexual discrimination, sixty years on, this language of selecting pupils for what is judged to be appropriate for them seems very dated compared, for example, to current talk about choice.

Some take the issue further with claims that cross-curricular and extra-curricular provision should be different for pupils of different ability, and that this is most easily achieved by separation. However, it is hard to imagine in practice what kinds of activities are appropriate for each without conjuring up unpalatable cultural snobbishness: should the grammar school offer orchestras and debating societies, while the secondary modern has football clubs and steel bands?

Of course many comprehensive schools operate some form of streaming to try to meet supposed different needs within the same school. With the requirements of the 1988 Education Act for all pupils to follow essentially the same National Curriculum, it could be argued that schools' freedom to offer this kind of differentiation is now quite limited.

Part of the argument for separation is that different teachers may be good at teaching different types of pupil, and this might be harder to address within a single comprehensive school. It may be that the demands of a grammar school to provide academic challenge for the highly able are sufficiently different from those of a secondary modern to provide a broad, practical and engaging curriculum for even the least able, that the required teaching skills are unlikely to be found within the same teachers.

2.1.2. Teaching is better targeted at narrow ability range

Related to the argument about the need for different curricula is the contention that, once one has decided what is most appropriately taught to a particular group of pupils, this is most effectively done in groups with a narrow spread of ability. Such targeting makes it easier to match curriculum content, pace and delivery to pupils' needs and readiness. The view that, at least for certain subjects, learning is best when pupils are grouped by ability seems to be widely held by teachers and others, as is evident from the setting that takes place within comprehensive schools. The fact that such an arrangement can easily be accommodated within a comprehensive school, however, somewhat undermines this as an argument for selection into different schools at 11. We may also note that despite widespread belief in the benefits of setting, it is not a view that is really supported by research evidence (Mosteller *et al*, 1996).

2.1.3. Grammar schools are meritocratic and socially redistributive by providing advantage for the bright but poor

It used to be common in England for Labour MPs and others from working class origins to attribute their success in life to the fact that they went to a grammar school. The grammar school was seen by many as offering the only way out of poverty and disadvantage for children from poor families. While grammar schools select by ability, the argument goes, in a non-selective system access to the best schools is available only to those who can afford to live in affluent areas or pay for private education. Hence the grammar school system weakens the link between income and educational success.

This argument was given a renewed currency recently with a report showing that social mobility has actually declined in Britain in recent years (Blanden *et al*, 2005) accompanied by press headlines such as "End of grammar schools has helped widen class divisions" (Daily Mail, 25 April, 2005). In fact, the report makes no mention of grammar schools; it is only in the associated press releases that the link is made. Moreover, it would be very hard to prove that the abolition of grammar schools was a cause of a reduction in social mobility, rather than just coincident with it.

2.1.4. Socially inclusive, as it keeps the middle classes in state education

One argument that has been put forward is that grammar schools represent an acceptable form of state education for the middle classes and so keep them out of the independent system. In this way, the existence of grammar schools creates greater social mixing than would be the case if they did not exist. Unfortunately, the current national pupil database for England does not include any pupils in independent schools, so we have not been able to test this claim.

The paper by Sullivan and Heath (2002), discussed in the next chapter, used data about children born in a particular week in March 1958 and identified some differences in social measures (including father's occupation and extent of parents' education) between those going to different types of schools including independent schools and direct grant schools (independent schools that received state funding for accepting pupils selected on the basis of the eleven-plus examination – a forerunner of the Assisted Places Scheme).

2.1.5. Academic elite should be a priority for education

This is an argument about values which maintains that it is more important for a small number of people to be highly educated than for education to be equally distributed. This kind of elitism is unusual in such bald form today, but arguments that emphasise economic needs for cutting edge science, medical research, technological development and advanced engineering may draw on it implicitly. It is arguable that from a purely economic point of view, the benefits of such elite contributions are greater than those of mass education, particularly if the associated costs are taken into account.

2.1.6. Grammar schools are a beacon of excellence

Related to the previous argument, but perhaps politically more acceptable, is an argument that it is important for our education system to have some schools that are simply excellent. Even if such schools are not available to everyone, simply knowing that they are there has a positive effect on the system as a whole. If truly excellent schools exist within a system, it would be morally wrong to close or transform them for the sake of egalitarian principles; a few excellent schools for a fortunate small group of pupils are better than no excellent schools at all.

Of course, this argument hinges on what we mean by excellent. How should we define 'excellence' and how would we identify it in practice? Of course, comprehensive schools could be just as likely to be excellent – though perhaps in different ways – so this is not necessarily an argument for selection. Nevertheless, it may well be the case that grammar schools are disproportionately likely to be seen as embodying this idea of an excellent school. It is notable, for example, that in a recently published list of the 11 secondary schools in England identified by Ofsted as 'outstanding' in three consecutive inspections, four are grammar schools (BBC, 2005).

2.1.7. Better academic results

One of the key arguments in favour of selection (sometimes the only thing considered) is that "selective systems" produce better academic performance. Most of the studies we review in Chapter 3 have set out to answer precisely the question of whether this is, in fact, true. The conclusions, as we shall see, are open to dispute. There are, however, a number of variants of this argument, that need to be examined in detail.

The first, rather simplistic, version of the argument is that in comparison with results achieved in comprehensive schools there is a clear advantage for pupils in grammar schools. In short, grammar schools get better results than comprehensives. One can see this, for example, in league tables of raw results where many of the top schools are grammar schools. The DfES league table for 2004, ordered by GCSE points per pupil shows 93 schools in the top 100 to

be grammar schools; of the top 50 only one, a City Technology College, is not a grammar school.

This comparison takes no account of the differences on entry between pupils in the two types of school, however the equivalent "value-added" league table of progress from Key Stage 2 to GCSE shows 70 of the schools in the top 100 to be grammar schools, with only 7 of these not in the top 100 schools in terms of raw GCSE points.

An obvious counter-argument is that it ignores what happens to the secondary modern (and technical school) pupils who fail to gain entry to the grammar schools and that comparison is needed instead between the achievements of the whole range of pupils across selective and non-selective systems. As discussed in Chapter 1, idealized selective systems of LAs comprising only grammar and secondary modern schools and non-selective systems of LAs with only comprehensive schools no longer exist in England, if ever they did. The lack of well-defined selective systems must be kept in mind when considering the published empirical evidence of such comparisons in Chapters 3 and 4, where we discuss methodological problems with those approaches.

The more sophisticated "value-added" argument comes with comparisons that attempt to compare like with like, by adjusting for the effects of initial differences between pupils at different types of schools. The majority of studies examined in Chapter 3 claim to provide fair comparisons by taking account of prior attainment and other factors but, as we shall find, there are difficulties, especially in obtaining appropriate measures.

Some feel it is so obviously true that high ability pupils will do better when surrounded by others of high ability, as in a grammar school environment, than they would in a less exalted environment, that no evidence is required to support such a claim. In fact, this is very far from being obvious, though establishing it empirically is quite problematic (see discussion of 'compositional effects' in Chapter 8).

Any attempt to compare the performance of different types of schools rests ultimately on value judgements. To compare performance across the whole range of academic achievement, unless one system is better in every respect, ultimately resort must be made to value judgements about the relative worth of different achievements. For example, if selective systems give some advantage to the most able, perhaps at the expense of those less able, as has been claimed by some, is this overall a benefit?

2.1.8. Selection operates elsewhere within the educational system

A final argument for selective secondary school education is that while selection by 11+ examination may be seen as controversial, other forms of selection operate within the English education system which are fundamentally similar but which seem to have escaped any real controversy. Two examples may be given here.

The first is the case of higher education, where entry to university has always been on the basis of academic selection. Students attend different institutions according to their abilities, which is a situation exactly analogous to selective education at age 11. The only difference is the age at which it happens. Why should it be acceptable for some universities to cream off the best students, but not acceptable for some schools to do the same?

The second example is the secondary 'comprehensive' school system in England in which a number of forms of implicit or explicit selection appear to operate. Specialist schools may select by 'aptitude', faith schools by religion and many schools effectively select by home postcode if they are over-subscribed. The overall effect of these different forms of selection may not be very different from the use of an explicit ability test, and may lose some of its advantages in terms of transparency, appropriateness of the selection criteria and social justice. We provide some evidence on the selectiveness of 'non-selective' schools in Section 7.2.

2.2. *The case against selection*

In this section we present counter-arguments against selection, i.e. in favour of an ideal comprehensive system. We begin with criticism of the selection tests themselves, which can never be ideal or without side-effects. There are also counter social and economic positions.

2.2.1. Selection tests are never fair or adequate

Grammar schools today and in the past have used a range of different tests for selecting their applicants at 11+. Inevitably some of these tests are better suited to the task than others, though it can be argued that even the best will never be adequate or fair. Evidence from Northern Ireland suggests that in practice the tests that have been used fall some way short of ideal in terms of their psychometric properties (Gardner and Cowan, 2005).

All tests are unreliable to some extent, so a person's score is partly a matter of chance. This means that for some, the decision to offer a grammar school place or not will be something of a lottery.

One way in which the adequacy of a selection test might be judged is in terms of its predictive validity. If a test at age 11 could accurately predict academic achievement at, say, age 16, then we might argue that such a test would be a good way of discriminating between those children at age 11 who were 'academic' and those who were not. The correlation between test scores at age 11 and achievement scores at age 16 is a measure of the validity of the prediction, a correlation of 1 indicating perfect prediction, and a correlation of 0 indicating no predictive ability at all. If the correlation was poor (close to 0) and a large number of those who 'failed' the test at 11 went on to achieve good academic results at 16, we might be less convinced that the test was really appropriate for the purpose of selection.

Fortunately, data exist for a number of different tests taken at age 11 and the corresponding performance at GCSE of the same pupils (e.g. Cognitive Abilities Test (CAT), MidYIS, London Reading Test (LRT). In none of these cases is the correlation much above about 0.7. Whether this is high enough to show adequate predictive validity is a matter of opinion.

Figure 1 illustrates how, with a correlation of 0.7 and a cut-off pass mark that selects 25% of 11 year olds, children can be wrongly selected or not selected. If, as a crude generalization, those 16 year olds who achieve in the top 25% are taken to be those who should have gone to a grammar school, we can see that about 78% go to the appropriate school for their ability, leaving around 22% wrongly allocated.

Figure 1: Mis/allocations of selection: an illustrative example

It is also the case that all tests contain bias. The intention may be to select on "ability" but the effects of, to give just a few examples, culture, home background, gender or personality are likely to affect scores on any test. This may be partly a question of how the construct of ability is conceptualised and implemented.

Even were the test itself completely unbiased, where individuals have a choice about whether to put themselves forward to take the test there is another form of bias. Some, perhaps those from lower socio-economic status backgrounds or from certain ethnic minorities, may be less inclined to apply to a grammar school. In Buckinghamshire, children may be "opted out" of the 11+ test by their parents, but more generally children are "opted in" to take an 11+ test by parents having taken the decision to apply to a grammar school.

2.2.2. *Ability is multi-dimensional and fluid*

A further set of objections to the use of a selection test to determine school entry relate to difficulties with the concept of "ability." The whole principle of selection depends on the claim that a single score on a test can capture all the different abilities relevant to predicting the whole spectrum of future academic achievement. This can never be strictly true, though a counter argument might be that it is adequate.

But there is, perhaps, an increasing acknowledgement by psychologists that ability is not uni-dimensional and that traditional conceptions of it are over simplistic. Gardner's (1983) theory of multiple intelligences is among the most commonly referred to (though perhaps also among the most commonly misrepresented) theories of ability. Other relevant and influential challenges

to the uni-dimensional idea include Sternberg's (1996) theory of successful intelligence. Both these theories would argue that, even if it were appropriate to allocate children to schools on the basis of a test, it would not be the kind of test that is commonly used.

However, the argument is more with the kinds of tests that are widely used, not necessarily with the principle of selection itself. The idea of specialist schools, now widespread in England, perhaps owes something to a multi-dimensional conception of ability – or 'aptitude', as it is being called in this context.

A further, and perhaps stronger, objection to the use of selection test is that ability, even if this means a good measure of an appropriate construct, is not fixed but changes over time (Moffitt *et al*, 1993). This is the nurture side of the nature/nurture argument that ability grows at different rates at different stages and to different extents for different individuals, all affected by the environments in which the development takes place. This fluidity probably accounts for much of the apparently wrong allocation of the population as illustrated in Figure 1. The test may have chosen the best 25% at 11+, but people change. Whereas, within a comprehensive school it is relatively easy to move pupils to different ability streams or teaching sets as ability changes, this is likely to be much harder for pupils sent to different schools.

2.2.3. Impact of failure on pupils not selected

An argument often put forward is that, for those who take a selection test and not offered a place at a grammar school, the experience of rejection and failure is damaging. The harm that this may cause includes damage to self-esteem, confidence and motivation to achieve. Rejected pupils may see themselves as failures and intrinsically worth less than those who succeeded.

Although this certainly seems plausible and is supported by substantial anecdotal evidence, we know of no systematic attempts to investigate whether real and lasting harm is done. No-one likes to fail and the experience may be unpleasant, but that is not the same as saying that real harm is done. It is possible, for example, that the experience of failure, though difficult at the time, is ultimately strengthening. However, in the absence of more rigorous evidence, this remains rather speculative.

It may also be pointed out that even without grammar school selection, the education system in England currently offers many opportunities for experiencing failure and the attendant risks of possible damage, for example the compulsory Key Stage tests. Whether these are worse for pupils in a secondary modern school than they are for those in the bottom streams of a non-selective comprehensive may well be questioned.

2.2.4. Adverse effect on primary school curriculum

If the 11+ selection test carries high stakes then it may be seen as inevitable in a selective system that primary schools will want to prepare their pupils for the test to give them the best chance of passing. Endless practising of 11+ tests might be a good way to improve Y6 pupils' chances of getting into the grammar school, but it certainly could not be described as a rich and interesting curriculum.

There is little evidence about the effects of selection on primary schools' curricula, though it has been argued that national assessments at Key Stage 2 (SATs) have had just this effect, quite independently of selection (Galton, 2002; Wiliam, 2001).

2.2.5. Socially divisive

This is the counter argument to Section 2.1.3 (p17). Irrespective of any effects on achievement, the way that selective systems separate children from those of different social classes into different schools may be seen as an argument against them. There does seem to be evidence that grammar schools contain children who are elite socially as well as academically (e.g. Gorard *et al*, 2003). Whether this is intrinsically a bad thing is open to debate. It could also be argued that this kind of social polarisation is not a necessary concomitant of selective education but just a contingent feature of the current mixed system; if there were more grammar schools, they might be more socially mixed.

It may be argued that even pupils attending the same comprehensive school are not fully integrated in practice. Understanding the full effects of any such social division is not really possible in the absence of evidence, and sadly there seems little available.

2.2.6. Selection compounds disadvantage

Perhaps a corollary of the divisiveness of academic selection is the argument that by deliberately creating schools in which pupils with relative social disadvantage are concentrated, their disadvantages are compounded. Moreover, separating these pupils from those whose parents are politically more empowered, their chances of gaining fair access to social and educational opportunity are reduced. In short, the system creates 'sink' schools in which the socially excluded are left to fester.

Though this argument may be sound, again we are limited by a lack of evidence by which to judge. It has been claimed that many of the remaining grammar schools are in, and draw their pupils from, areas of relative social advantage in England (Atkinson *et al*, 2004; 2006). Comprehensive schools at the bottom of the league tables are more likely to be in socially disadvantaged areas of the country, rather than in areas creamed by grammar schools. Given the levels of social segregation by residence in this country, and the tendency of schools to recruit from a particular area, it is not likely to be academic selection so much as "postcode" selection at blame for the phenomenon of sink schools.

2.2.7. Socially disadvantaged should be a priority for education

The counter-argument to the contention that the elite should be the priority for education (see Section 2.1.5, p18) is the view that it is those who are most disadvantaged who should receive priority treatment. This argument may be supported by economic reasoning that counts the costs of crime and other forms of social exclusion that might be reduced if better education were provided for the most disadvantaged. Of course, it can also be argued on grounds of natural justice and fair play.

There is no reason in principle why one should not wholly support these arguments for the need for education to redress social inequalities and yet

believe that a selective system provides the best way to do so. If, for example, priority is reflected in funding then one could simply give more money to schools with the most challenging populations.

2.2.8. Selection limits parental choice

Finally, we note an argument that points to the ideal of parental (and pupil) choice of which secondary school to go to, perhaps as a necessary condition for market forces to operate. In a selective system parents can apply to a grammar school, but their son or daughter will not necessarily be offered a place. Any remaining choice is then limited to non-selective schools in the area. Whether this is a bad thing, though, may be a matter of judgement.

Here again, we must compare the alleged disadvantage of selection with the reality of comprehensive organization. Within the current system in non-selective areas there is often perceived to be very little real choice anyway over which school a child can attend. What choice there is may be purchased only by those who can afford to live in affluent areas or provide their own transport to get to school. However, Burgess, Propper and Wilson (2005) argued that more parental choice combined with flexibility in the supply of school places will reduce sorting of students by income and ability, and competitive pressure will urge schools to improve standards.

2.2.9. Worse academic results

This is the contrary position to the argument in Section 2.1.7 (p18) and essentially takes the same form. Where it is supported by empirical evidence, the emphasis is likely to be on analyses that look at the whole spectrum of schools, including the performance of secondary moderns as well as the grammar schools, and on value-added analyses that take account of differences in the ability of pupils at different types of school on intake. A review of this evidence is presented in Chapter 3, together with discussion of methodological challenges to interpreting it (Chapter 4).

A variation on the simple focus on the average achievements of pupils in one system or other is an argument that looks at how the range of achievement may differ. For example, it could be argued that although pupils may achieve the same on average, within a selective system the range of achievements is greater, with a larger gap between the performance of the best and worst. It could also be argued that the relationship between a factor such as social class and achievement might be stronger within a selective system. In other words, even if pupils as a whole do not do better in non-selective systems, pupils from backgrounds with greater disadvantage are helped, relatively speaking, compared with what would be expected in a selective system.

PART II
LITERATURE REVIEW

3. Existing Empirical Evidence on the Effects of Selective Systems

In this chapter we review the literature on the major studies that have been carried out to compare pupils' attainment in selective and comprehensive school systems. We have considered only those studies in which the researchers have drawn their conclusions from an analysis of empirical data. Such studies have occurred from about 1980 onwards and we present our review in the chronological order in which the research was published.

For each of the research studies we have organized our review in the same way. This is a brief overview including why the research was conducted and for whom, the nature of the data collected and the methodology by which it was analysed and the conclusions the authors came to. Further descriptive detail of each research study is included where appropriate. We include critical comment by other researchers where it has been made on some of the studies and also offer our own critical appraisal of each piece of research. An overview of the problems encountered in empirical research is given in Chapter 4, and, in the light of this, a summary of the literature review is given in Chapter 5.

Our main focus has been on research conducted in England, but at the end of this chapter we have included a review of some of the research concerning selection and secondary education in other countries.

3.1. Major studies in the 1980s

3.1.1. Steedman (1980): Progress in Secondary Schools

This research was commissioned by the Department for Education and Science to evaluate aspects of progress in selective and non-selective schools in England.

Data

The data used were from the National Child Development Study (NCDS) of individuals born in a particular week in March 1958. The study used data related to pupils' experiences and educational performance made available from the 1974 sweep, when its members were sixteen years old. The pupils were in secondary schools between 1969 and 1974 during which time many comprehensive schools were newly established. Steedman limited her study to those pupils who had spent all their secondary years in a school whose status did not change, to attempt a fair comparison of those who attended grammar and secondary modern schools with those who attended comprehensive schools.

Methodology

Multiple regression with a range of explanatory and outcome variables. The main explanatory variable was type of school with many other variables controlled for. She defined 19 outcome variables (the explanatory and outcome variables are delineated below).

Results and conclusions

Steedman found no significant differences in the progress between pupils in selective systems (i.e. those who had attended a grammar school or secondary modern) and those who had attended a comprehensive school once allowance had been made for the relatively lower entry attainments and social class background of pupils who went to comprehensive schools.

Student attitudes were found to be good in secondary modern schools.

Behaviour in comprehensive schools was perceived, by teachers, to be worse.

Those who had attended grammar school showed higher academic attainment, but on a value added measure between 11 and 16 there was no difference.

Grammar school pupils generally had higher expectations of staying on at school and entering the professions after higher education, than leaving and entering manual or clerical jobs. They also gave a high rating to helping others as an important aspect of a job.

Further detail

Explanatory variables

School type and age range, social composition, single sex or mixed, region (north or south of England), sex of the pupil, parents interest in school, ability test score at 11 years old, the age at which the pupils' parents left school, whether a pupil's class for English and mathematics was streamed, mixed ability or remedial

Outcome variables

Basic skills attainment (reading and maths), self-rated performance in reading and maths, truancy (from school teacher report), proportions stayed off school, school behaviour rated by teachers (Rutter school scale total), liking of school (pupils' view on 'I don't like school'), likely age of leaving school, proportions wanting part study with job, leaving school at 16, proportions wanting advanced course of study after leaving school, proportions wanting profession/manual/clerical first job, proportions considering 'being well paid'/'chance of promotion'/'opportunity helping others' as important in a job; proportions of parents' satisfied with child's education in present school, and behaviour ratings by parents.

Results

Basic skills (maths and reading tests and self-ratings): on average, children attending a grammar school scored highest while those attending a secondary modern school scored a little lower than those attending a comprehensive school. However, there was no evidence that pupils in comprehensive schools had made any worse progress than those in the

combined system of grammar and secondary modern schools. A similar picture was found for self-ratings, but in this case, grammar school pupils were found to rate themselves lower in mathematics and reading ability than pupils in comprehensive and secondary modern schools; the latter two being similar.

In terms of a relationship with prior attainment at 11 years old, pupils who were in the top attainment group at age 11 before the start of secondary school did as well in comprehensive schools as in grammar schools.

With regard to *truancy and general behaviour*, based on the views of teachers, pupils in comprehensive schools seemed to have a higher rate of truancy, regardless of their social class. Steedman stated that this could be interpreted as higher incidence of truancy or higher rate of detection of truancy. In grammar schools, lower ability pupils were more likely to truant and there were also social class differences. Teachers of pupils in comprehensive schools also regarded their pupils as particularly badly behaved compared to teachers in selective systems (grammar and secondary modern school) as a whole.

Findings related to *liking for school* were inconsistent with findings related to truancy and general behaviour. Pupils in comprehensive schools were no different from grammar school pupils in liking for school, and secondary modern pupils were most likely to indicate a liking for school. No differences were found between comprehensive and selective systems of grammar and modern schools combined.

For the outcome variable *likely leaving age* pupils in grammar schools tended to anticipate later leaving, but overall, the average planned leaving age of pupils in comprehensive schools was slightly later than that intended by the pupils of the grammar and modern schools combined.

Some differences were found between the three types of schools in relation to *further study and preferred first job*. Grammar school pupils were found most likely to want advanced courses after leaving school and professional status in their first job and least likely to want manual or clerical first jobs. Secondary modern school pupils ranked the opposite from grammar school pupils while comprehensive and secondary modern school pupils were similar. However a comparison between comprehensive schools and grammar and modern schools combined, showed there were no significant differences between the two systems.

For the outcome variable *pupils' view on what is important in a job*, overall, there were no statistically significant differences between the school types relating to proportions considering being well paid or chance of promotion as important in a job. Grammar school pupils were highest in considering opportunity of helping others as important in a job and comprehensive and secondary modern were similarly low in this measure. However, no differences were found when compared between the two systems of comprehensive and grammar and modern schools combined.

For the outcome variable *parents' views*, a higher proportion of parents whose children attended grammar and secondary modern schools were satisfied with their children's education in their present schools, compared to parents with children in comprehensive schools. There were no differences in terms of parents' rating of their children's behaviour.

3.1.2. *Steedman (1983): Examination Results in Selective and Nonselective Schools*

Steedman's follow up study investigates examination results of the pupils in the NCDS study (1980), after they had taken public examinations at age 16, and possibly age 18. The main question being addressed in this study was 'Did selection for secondary school make a difference?' Steedman divided the data roughly into quintiles using average test scores at 11 years old and subdivided each of these by social class, defined by father's occupation being manual or non-manual.

Table 2: Mean number of O' level equivalent passes achieved by pupils in selective and non-selective systems, split by prior ability and social class.

Quintile of test scores at 11		Highest				Middle				Lowest	
Social class		GS+SM	Comp	GS+SM	Comp	GS+SM	Comp	GS+SM	Comp	GS+SM	Comp
Non-manual Pupils	Mean number O'level passes	6.4	5.2	3.7	3.3	1.7	2.1	0.6	0.7	0.1	0.4
	Number of pupils	317	63	191	44	108	54	66	38	35	27
Manual Pupils	Mean number O'level passes	4.7	4.8	2.7	2.7	1.0	0.9	0.3	0.4	0.1	0.2
	Number of pupils	217	64	246	126	266	143	234	169	280	208

The table above is derived from Steedman's report. It shows the number of pupils in each category and the mean number of O-level equivalents obtained by them up to 1974. O' level equivalents were a grade of A, B or C at GCE O' level or a pass at CSE grade 1. Steedman also gave data for secondary modern and grammar schools separately, but for this report we have put the emphasis on the grammar school – secondary modern school combination. However, we note the relatively small numbers in the data set which limits any generalisation of the results. Also of the 747 pupils in the study who had attended a grammar school, 699 (94%) were in the top two quintiles as measured by ability at age 11, so the lowest three quintiles are essentially a comparison of secondary modern and comprehensive pupils.

Steedman observed that within each quintile and social class, grammar schools alone showed the best results and secondary moderns the worst. The grammar-modern combination results are better than those of comprehensive pupils for the more able children for 'non-manual' pupils and similar for 'manual' pupils.

Steedman says that a similar picture was also found for the average number of A-levels. The mean number of A-level passes for the pupils in each category at age 11 are shown in the table.

Table 3: Mean number of A Level passes for pupils in each type of school, split by social class and prior ability

Quintile	Highest				middle				Lowest	
Social class	GS+SM	Comp	GS+SM	Comp	GS+SM	Comp	GS+SM	Comp	GS+SM	Comp
Non manual	1.5	1.2	0.5	0.5	0.1	0.2	0	0	0	0
Manual	0.9	0.6	0.2	0.2	0.1	0.1	0	0	0	0

We note that at A-level the actual numbers of pupils taking the examinations are not known but will be much smaller than the number of pupils in each quintile category, making any comparison rather meaningless.

Limitations to Steedman's studies

Steedman was clear that due to limitations on the evidence, caution should be adopted when discussing the results of her studies. The highest-scoring 2% of pupils at age 11 had all gone to grammar schools and could not be compared with other pupils. She further highlighted that neither study could give an answer to whether one system is better for children than the other as the data did not come from complete selective and comprehensive systems. The pupils in the sample attended secondary education from 1969 to 1974, during which time piecemeal changeover to comprehensive schooling occurred. The average intakes for comprehensive schools were most like those for secondary modern schools. Sample pupils from comprehensive schools scored lower on tests of attainment or ability at 11 years old, and were more likely to be 'working class' than the average for the combination of grammar and secondary modern schools.

Steedman concluded that it did not appear, from the evidence she had available, that comprehensive pupils were doing any better or worse than pupils in the combination of grammar and secondary modern schools. She did note several reservations about using examination results as basis of comparing large numbers of pupils, due to differences in subjects and aggregating over subjects. However, she noted at the start of her study

> …schools called comprehensive schools may not generalise to present day comprehensives; they (her study) are observations of how pupils were faring in schools which were not true comprehensives but which coexisted with selective schools.

We note that in 1983 Steedman was raising the question of what is a true comprehensive school, a problem that is still part of the selection issue in 2005.

Criticism of Steedman by others

Many researchers have commented on the use of the NCDS data. In a review of Steedman's study, Fogelman (1984) argued that Steedman's sample was limited to all children in England born in a given week in 1958, and therefore could not be systematically related to the type of secondary school attended. He also highlighted the problem of sample attrition. In another review of Steedman's studies, Marks & Cox (1984) pointed out that her analyses refer to the effects of schooling during a relatively early stage of the change to

comprehensive schools (the pupils reached the age of 11 in 1969 and reached the age of 16 in 1974). Furthermore, the actual numbers used in most analyses were considerably less than the original NCDS cohort as the study did not make use of sample from Scotland and Wales, or those who went to independent schools, or those who changed secondary school or those who lacked information on a number of background variables. In many analyses, it only involved 4056 pupils – 912 in grammar schools, 1765 in secondary modern schools and 1379 in comprehensive schools. The sample used in the 1983 report was even smaller as examination results could not be traced for all the pupils concerned. The maximum size of the sample was 2896 pupils of which 747 were in grammar schools, 1213 were in secondary modern schools and 936 were in comprehensive schools. In both cases, the sample may not be fully representative. What is more, since the children in the sample were born in the same week in March, they were in the middle of their school year group, and their experiences may not be fully representative of the whole year group.

Marks and Cox also commented on the use of the reading test at the age of 16 as a test to measure attainment in Steedman's study. The test was a sentence completion test and has been criticised as showing a significant ceiling effect for 16-year-olds. This ceiling effect may have contributed to the findings that there was little difference in reading at 16 between pupils from grammar schools and secondary modern schools combined, compared with pupils from comprehensive schools. Marks and Cox argued that Steedman's studies lacked any analysis of possible correlations between the input and output variables. Most of her analyses used multiple regression and the results presented were the residual scaled differences remaining after adjustments had been made for background variables. They claimed that Steedman actually found results that were similar to the ones they (Marks, Cox and Pomian-Srzednicki, 1983) had found in their 1983 study (the discussion on this study will follow). Steedman found that especially after adjustment, pupils at grammar and secondary modern schools performed better than pupils at comprehensive schools, but only one of these differences was statistically significant. It is possible that the differences would have been statistically significant if the total sample had been larger.

To these criticisms we may add a general one of our own, that the use of baseline matching into broad categories such as quintile groups is likely to overestimate the differences between achievements of grammar and non-grammar school pupils. For more detailed discussion of this issue, see Section 4.1.2 (p108).

It is also notable that pupils at the secondary modern schools indicated a liking for their schools. Given this was 1969 to 1974 they were probably following a vocationally orientated curriculum. Whether they would have said the same had they been following a more academic curriculum is an interesting question of speculation.

3.1.3. Marks, Cox and Pomian-Srzednicki (1983): Standards in English Schools: Report No 1

This study was published by the National Council for Educational Standards, an independently funded body who were concerned about the possible lowering of standards due to the introduction of comprehensive schools.

One of the major aims of the study was to examine the effects of selectivity on examination results. The authors accepted that there is more to education than just examination results but argued that they are important in providing impartial indicators of school's educational performance. The authors noted that between about 1960 and 1980 the proportion of secondary pupils attending comprehensive schools rose from less than 10% to over 90%. They noted that the policy of comprehensivisation had been called by some as "an experiment with the lives of millions of children " and so they set out to analyse the result of this "massive change", noting a lack of any other monitoring particularly by the DES. They stated that their results indicated that examination results for the selective sector are considerably better than those for comprehensive schools, and that both grammar schools and secondary modern schools, in their different ways (authors' phrase), are enabling their pupils to gain good examination results. They then suggested (authors' word) that diversity in our school system may be more desirable than homogeneity and schools that concentrate their efforts on specific groups of pupils may be more successful than those which attempt within one institution to cater for the needs of a whole school population.

Data

Examination results of schools in England in 1981, including O-level, CSE and A-level results, obtained from 54 of the 96 Local Authorities in England and Wales at the time (2,100 schools and more than 350,000 fifth-year pupils). Results were presented as passes per pupil and a points score per pupil for all pupils in a particular cohort.

Selectivity was measured by the percentage of pupils who in 1976 entered a grammar school out of the total LA cohort who entered secondary school.

The vast majority of those who sat examinations in 1981 would have entered secondary school in 1976 at age 11. Allowance was made for those pupils who entered secondary education at a different age.

Explanatory variables

Social class, type of school, level of expenditure in the LA, pupil to teacher ratio, proportion of pupils of non white origin or born abroad

Output variables

The number of O-level and CSE grade 1 passes per pupil (denoted OC1).

The number of CSE passes at grades 2 to 5 per pupil.

The average number of points per pupil. (7 points for O-level grade A; 5 points for O-level grade C or CSE grade 1 assumed equivalent; 1 point for CSE grade 5)

The number of A-level passes

The average number of A' level points per pupil (5 points for grade A and 1 point for grade E)

Methodology

Correlation analyses between aggregated examination pass rates for each LA in the study and each of the explanatory variables except social class. Multiple regression using the explanatory and output variables as above.

Results and conclusions

Higher examination results tended to be obtained in LAs which had a higher proportion of pupils in schools which select by ability.

Examination results at A-level, O-level and CSE grade 1 in comprehensive schools were consistently below the national average.

Pupils from comprehensive schools performed better than those from secondary modern schools, although the latter did better than expected.

It was estimated that the number of examination passes in a fully selective system of grammar and secondary modern schools would be between 30% and 40% higher than in a fully comprehensive system.

More detail

A quota sampling method was used to ensure representativeness of Local Authorities (geographical distribution, urban and rural areas, and social class). Counties were included from the North, the Midlands, East Anglia, the West Country, the South and the South-East, as were LAs from all the Metropolitan Districts – Greater Manchester, Merseyside, Tyne and Wear, West and South Yorkshire, the West Midlands and London. Three LAs in Wales were involved.

Each of the five output variables was analysed for all O-level and CSE subjects together and then for what the authors called the eight main subjects, (mathematics, physics, chemistry and biology as science subjects; English, French, history and geography as arts subjects). Mathematics and English were also considered separately. A-level results were analysed for the four main science subjects and arts subjects taken together.

Some limitations of the use of A-level results were highlighted. These included variations in staying on rates in different areas, the absence of results from further education colleges or tertiary colleges, the considerable transfer that takes place at 16+ both into and out of school sixth forms and the growing tendency for schools to share their sixth form provision. Therefore, although the authors made analyses relating to A-level results, their limitations were recognised.

Correlation analysis was done between examination results and selectivity. Results showed that selectivity correlated consistently with examination performance. All correlation coefficients between selectivity and examination results were positive, with values ranging from 0.36 (for O/C points per pupil over all subjects) to 0.52 (for OC1 passes per pupil in the 8 main subjects). This indicates that higher examination results tended to be obtained in LAs which had a higher proportion of their pupils in schools which select by ability.

Examination performance was found to be negatively correlated with LA expenditure, indicating that the more an LA spent the poorer were the results. There were no clear results from correlation analyses between examination performance and pupil to teacher ratios or with proportion of non-white pupils or pupils born abroad,

The next analysis was a comparison of the overall attainments of the pupils in comprehensive, grammar and secondary modern schools. Examination performance at CSE, GCE O-level and A-level were compared against national averages for England. These averages were calculated based on 1,897 local authority schools with 349,000 fifth-year pupils. The main purpose of

presenting these national averages was to provide norms or benchmarks as a basis of comparison. Table 4 shows a summary of the national examination results for England in 1981, and for the three types of secondary school. For example for OC1 passes per pupil, the national average for England is given as 2.14, for comprehensive schools as 1.91, for secondary modern schools as 1.39 and for grammar schools as 5.51.

The results from comprehensive schools across all three examinations were consistently somewhat below the national average. The authors believed that the lower results might have been due to including areas where comprehensive schools existed along side grammar schools, and thus might have been creamed of top ability pupils. They therefore carried out a further analysis looking exclusively at schools that were fully comprehensive (17 LAs, 330 schools, 72,000 fifth-year pupils). The results found were identical to the ones they found with all comprehensive schools. The authors therefore argued their sample of comprehensive schools is representative of the performance to be expected from a wholly comprehensive system of schools.

Pupils in secondary modern schools obtained O-level results that were only a little lower than those achieved by pupils in comprehensive schools, and their CSE results were considerably higher than the national average. Their A-levels result were very low but secondary modern schools had few pupils with A-levels.

Pupils at grammar schools obtained high results at O-level and A-level examinations, but relatively low in their CSE results, which the authors considered to be due to the fact that many grammar schools did not enter pupils for CSE examinations.

Table 4: Summary of OC1 passes, O/C and A-Level for comprehensive, secondary modern and grammar schools in England in 1981, extracted from Marks et al (1983)

	All subjects	Main subjects	Maths	English
OC1 passes per pupil				
England	2.14	1.53	0.27	0.50
Comprehensive	1.91	1.34	0.23	0.45
Secondary Modern	1.39	0.89	0.17	0.32
Grammar School	5.51	4.44	0.72	1.35
O/C points per pupil				
England	22.7	15.4	2.8	5.1
Comprehensive	21.4	14.4	2.6	4.8
Secondary Modern	19.4	12.3	2.3	4.5
Grammar School	39.0	31.2	5.1	9.1
A-Level passes per pupil				Science
England	0.34	0.23	Art Subjects	Subjects
Comprehensive	0.30	0.21	0.10	0.13
Secondary Modern	0.02	0.01	0.08	0.13
Grammar School	1.29	0.95	0.01	0.00
			0.41	0.54
A-Level points per pupil			Art Subjects	Science
England	0.93	0.63	0.26	Subjects
Comprehensive	0.76	0.53	0.20	0.37
Secondary Modern	0.04	0.04	0.03	0.34
Grammar School	3.69	2.75	1.15	0.01
				1.60

OC1 pass = O-level passes at grades A,B and C together with CSE passes at grade 1
O/C= the average number of points for O-level and CSE together.

One of the objectives of the authors was to evaluate the result of comprehensivisation. However they recognised that their data on grammar schools and secondary modern schools was over weighted with grammar schools and thus did not fairly represent a fully selective system. In order to make valid comparison of the examination results between the school systems (fully selective and fully comprehensive), the authors made adjustments to the data. Estimates of the averages for pupils in a fully selective school system were calculated. This was done by combining the average values of grammar school and secondary modern school based on the information published by the DES about the proportions of pupils attending different kinds of schools from year to year. The data show that when fully selective systems of schools were common in this country, the proportion of pupils attending grammar schools was generally somewhere between 25% and 30%. Two estimates of the averages were calculated. The lower estimate (S1) was calculated by assuming that 25% of pupils attend grammar schools and 75% attend secondary modern schools (S1=0.25G+0.75M), while the high estimate (S2) was calculated by assuming that 30% of pupils attend grammar schools and 70% attend secondary modern schools (S2=0.30G+0.70M). Average examination performance for comprehensive schools was also adjusted as there were 'restricted range' comprehensive schools (in areas where grammar school exist) that had very few or no pupils from the top 20% of the ability range. Estimation was done to exclude results based on the 'restricted range' comprehensive schools. The estimated values of the fully selective school system (S1 and S2) were than compared with the actual values of the comprehensive school system (C) and

with the estimated values of the fully comprehensive school system (CF) which excluded restricted range comprehensive schools, resulting in the ratios S1/C, S2/C, S1/CF and S2/CF shown in Table 5.

Table 5: Comparison between selective and comprehensive schools

	OC1 pass / pupil				O/C points / pupil				A-level			
	All	Main	English	Maths	All	Main	English	Maths	Pass all	Points all	Pass main	Points main
S1/C	1.27	1.33	1.29	1.35	1.14	1.18	1.18	1.15	1.13	1.25	1.14	1.30
S2/C	1.38	1.46	1.40	1.48	1.18	1.25	1.23	1.21	1.33	1.50	1.33	1.55
S1/CF	1.19	1.24	1.21	1.29	1.11	1.14	1.16	1.12				
S2/CF	1.30	1.36	1.31	1.42	1.30	1.21	1.21	1.18				

The authors noted the results of these calculations show consistently that substantially higher O-level, CSE and A-level examination results are to be expected for pupils in a fully selective system than in a fully comprehensive system and this finding applies to all the indices of examination success that were compared. The differences are higher for OC1 passes per pupil, ranging from 27% to 35% for S1 and from 38% to 48% for S2. For O/C points per pupil the differences range from 14% to 18% for S1 and from 18% to 25% for S2. For A-level passes per pupil, the national differences averaged 14% for S1 and 33% for S2. Although the differences were not very high, the authors still considered them appreciable.

Examination results were also compared between comprehensive and selective systems of schools by social class. The authors cited other research as showing that social class exerted great influence on educational attainment. From their own data, the authors found the strongest correlation was between attainment and social classes 4 and 5 (occupation of head of household as semi skilled or unskilled manual worker), so they used these classes to group the LAs into three categories, A, B and C, using census data. Group B had pupils who on average were close to the national average, with Group A with pupils below the national average and Group C above. The results are shown in Table 6.

Table 6: Comparison between selective and comprehensive schools by social class

	OC1 pass / pupil				O/C points / pupil				A-level			
	All	Main	English	Maths	All	Main	English	Maths	Pass all	Points all	Pass main	Points main
Group A												
S1/C	1.19	1.20	1.13	1.14	1.09	1.09	1.09	1.12	1.05	1.13	1.15	1.18
S2/C	1.29	1.31	1.24	1.24	1.13	1.15	1.14	1.17	1.24	1.35	1.35	1.40
Group B												
S1/C	1.34	1.45	1.44	1.45	1.17	1.25	1.24	1.25	1.11	1.21	1.16	1.27
S2/C	1.46	1.58	1.56	1.59	1.22	1.31	1.29	1.30	1.32	1.44	1.37	1.51
Group C												
S1/C	1.04	1.06	1.02	1.00	0.99	0.99	1.01	0.93	1.15	1.27	1.11	1.27
S2/C	1.14	1.18	1.12	1.10	1.03	1.05	1.06	0.98	1.37	1.51	1.33	1.50

When comparing examination attainment within these groups, results similar to the whole dataset were found. For Groups A and B, there were substantial differences between results estimated for pupils in a fully selective system compared with a fully comprehensive system in favour of the selective system with smaller differences in Group C. For example, in Group B, OC1 passes per pupil ranged from 34% to 45% higher for S1 and from 46% to 59% higher for S2; the O/C points per pupil ranged from 17% to 25% higher for S1 and from 22% to 31% for S2. For Group C, OC1 passes ranged from 0% to 6% higher for S1 and from 10% to 18% higher for S2; the O/C points showed no differences for S1 and 3% to 6% higher S2. At A-level the authors noted that points per pupil were in all cases greater than for passes per pupil. The differences between social class groups were similar but were a little higher for groups B and C than for group A.

Conclusion

The authors came to the following conclusion, writing it in capital letters in their report to emphasise it.

> The results presented in this section (on selectivity and social class) and the previous section (on relationships between factors such as social class, expenditure, pupil teacher ratios, selectivity and ethnicity) strongly indicate that substantially higher O-level, CSE and A-level examination results are to be expected for pupils in a fully selective system of schools compared with pupils in a fully comprehensive system of schools. This finding applies to all the indices of examination success which we studied and, according to our data, is as robust as the generally accepted finding that examination results are correlated with social class.

Criticism of Marks, Cox and Pomian-Szrdnicki

Fogelman (1984), in his review of the problems involved in comparing examination performance in selective and comprehensive schools, noted, unlike Steedman, Marks *et al* did not identify and exclude the pupils of transitional schools and those who had not been at the same school throughout in their analyses. This was mainly because the data Marks *et al*

had did not allow them to do so as no information was given relating to the status of the school. However, Fogelman argued that these pupils might have experienced both types of schools, and the results derived from such analyses can be questioned. Criticisms were also made of Marks *et al*'s analyses in allowing for social class differences. Marks *et al* compared the examination results of pupils of different types of school within groups of LAs which they claimed to be socially homogeneous. Fogelman argued that the DES social class measure used by Marks *et al* is extremely crude for this purpose, identifying only three 'clusters' of local authorities, based on the proportions within them of social classes IV (semi-skilled) and V (unskilled manual) groups. He claimed that others have pointed out that this is hardly likely to be sensitive to the social differences between authorities. His point was illustrated by analysing the proportions of children who went to selective or comprehensive schools, whose fathers were in semi- or unskilled manual occupations within each of the three groups of local authorities using the NCDS data.

Table 7: Percentage in social class IV and V, by sector, for three local authority groups (NCDS unpublished data), (Fogelman, 1984, p. 39)

	Selective	Comprehensive
Cluster A	12.8	17.9
Cluster B	17.6	20.3
Cluster C	21.3	22.5

From this illustration, Fogelman pointed out that there were more pupils from social class IV and V who went to comprehensive schools than to grammar schools in cluster A and B. There was not much difference in cluster C. He further compared this finding with Marks *et al*'s finding. Marks *et al* (1983) had found significant benefit of the selective system for cluster A and B but the contrast was relatively small within cluster C. Fogelman suggested that the greater contrast that they found within cluster A and B might simply be the result of inadequate allowance for social class.

Crook *et al* (1999) noted that Gray, Jesson and Jones (1984) analysed the same NCES data as Marks *et al* and found no reason to conclude that selective systems produce superior examination results. These authors also claimed Marks *et al* had not allowed adequately for social class and other related factors. In their own analysis they controlled for these factors, and reached a different conclusion to Marks *et al*.

We note that Table 4 shows Grammar schools performing better than other schools as might be expected given their intake. We agree with the authors that a comparison with a comprehensive system should include both grammar schools and secondary modern schools. These authors however, encountered the problem of finding fully selective and fully comprehensive areas in England, which they could compare. Thus they made adjustments to the data based on 25% and 30% selection levels to create such systems and estimate the effects. Their results show better results coming from a fully selective area, but the details of the calculations and the numbers of pupils concerned in various categories have not been shown nor the sensitivity to the percentages chosen and it must be stressed that these are estimates, not real data. It is notable that in 2005, the national percentage attending grammar schools is about 4% but with wide variation across those LAs that

still retain at least one grammar school. It is also notable that in several places in their report these authors highlight how well the secondary modern schools are doing, as if to emphasise the merits of the selective system. Although this may have been the case in 1981, we note that this was before GCSE examinations replaced O-level and CSE examinations and the introduction of the National Curriculum in 1988, and so the relevance to modern day is questionable.

3.1.4. Marks and Pomian-Srzednicki (1985): Standards in English Schools: Report No 2

The National Council for Educational Standards (NCES) published a second report by Marks and Pomian-Srzednicki on examination results in secondary schools in England. In this report, they carried out a similar analysis of the examination results for 1982 as they had done in the previous report for the 1981 examination results. They compared the results of their analyses for the two years, and found them to be very similar. They acknowledged criticism of their previous report and that they were aware of its limitations and other interpretations of the data. However, the authors stressed their intention was again to raise questions concerning education policy and the comprehensive experiment, particularly whether this was the best way of providing a quality education for all children. They again concluded that a selective system of grammar schools and secondary modern schools would lead to higher achievement than a comprehensive system.

The data and methodology were similar to their previous report, so we will not give so much detail in this update.

The data were obtained from 57 LAs, 49 of whom had been in the previous study. A total 380,000 pupils from 2,175 schools (including schools in Wales and independent schools) were involved in this study. The main analyses again were concentrated on state-maintained schools in England. The authors again reported that great care had been taken to ensure that the sample was representative and so included LAs from all parts of England covering inner city, suburban and rural areas. They said this also ensured an adequate numbers of different types of schools.

The authors also obtained more data on social class. They acknowledged the criticism re social class of their previous report but maintained that groups 4 and 5 were the most highly correlated social class measure with their indicators of O-level performance and used these again in their main analysis, noting the DES had done likewise in its statistical bulletin 16/83. They noted that social class is both difficult to define and measure and that it can only be taken into account in a broad and crude sense, but said they were using the best measures available. They extended these to social classes 1, 2 and 3 but said they made no major differences to their results.

In general, the results of the second study were similar to the results of the previous year. The relationship between selectivity and examination results was again examined using the 1977 pupil cohort, this being the year when pupils who sat examinations in 1982 would mostly have entered secondary school. The selectivity index varied from zero (LAs where there were no grammar schools) to 0.31 (the national average of the sample was 0.05). Correlation coefficients between selectivity index and examination measures ranged from 0.39 (OC points per pupil, for all subjects) to 0.53 (OC1 passes

per pupil for the main subjects) (the main subjects and codes used for the aggregated results were the same as for the previous report; see Table 4).

Examination results of the different types of school were compared with the national average. Table 8 below is similar to Table 4 above and is included to illustrate the similarity of the results obtained in both 1981 and 1982.

For comprehensive schools, OC1 passes and O/C points per pupil were consistently somewhat below the national averages for all LA schools, but this is notably less so for the O/C points score compared to the OC1 passes per pupil. However CSE passes at grade 2 to 5 for comprehensive schools at 3.51 passes per pupil was a little above the national average of 3.34 and similarly for points per pupil at 11.5 and 10.8 respectively. These differences exist for all four indicators used by the authors; that is for all subjects, the 8 main subjects and mathematics and English. Pupils at comprehensive schools in LAs that were fully comprehensive obtained, on average, results that were about 5% below those for pupils in comprehensive schools in LAs where there was at least one grammar school and thus some possible creaming of the comprehensive schools.

Table 8: Summary of OC1 passes, O/C and A-Level for comprehensive, secondary modern and grammar schools in England in 1982, extracted from Marks et al (1983)

	All subjects	Main subjects	Maths	English
OC1 passes per pupil				
England	2.13	1.59	0.27	0.32
Comprehensive	1.98	1.47	0.25	0.30
Secondary Modern	1.47	0.98	0.18	0.24
Grammar School	5.54	4.70	0.77	0.75
OC points per pupil				
England	22.5	16.2	2.8	3.4
Comprehensive	22.3	15.9	2.7	3.3
Secondary Modern	20.8	13.6	2.4	3.3
Grammar School	39.0	32.8	5.3	5.0

For secondary modern schools the authors noted that examination results per pupil were consistently good, particularly in mathematics and English. They noted further that since O-levels were originally intended for the top 20% of the ability range, secondary modern schools would not be expected to obtain many OC1 passes, yet the OC1 passes per pupil which they obtained are only a little below those for comprehensive schools, while for O/C points per pupil there was very little difference. Examination results per pupil for grammar schools were as expected, consistently high.

The authors again used their 25% and 30% selectivity assumptions to estimate what the results of a fully selective system might look like. They performed a number of analyses, using both national cohort data and within social class groups and these indicated that examination results per pupil would be substantially higher for a system of selective schools (secondary modern and grammar schools) than for a system of comprehensive schools; OC1 passes per pupil were estimated at between about 30% and 40% higher and O/C

points were estimated at between about 15% and 25% higher. They found similar differences existed for A-level results.

The authors concluded that this second NCES report increased confidence in the reliability of the research techniques and the validity of the data but they did note the need to be tentative about the inferences. However, they did conclude that as well as social class, selectivity was an important factor affecting examination performance. They stated that pupils in selective schools still seem (authors' word) generally to attain more and better passes than those in comprehensives.

They again emphasised their principal conclusion.

> The weight of evidence supporting our conclusion-that, with the exception of passes at CSE grades 2-5, substantially higher examination results are to be expected for pupils in a fully selective system of schools compared with pupils in a fully comprehensive system-is therefore considerably greater than when our previous report appeared and it remains valid even when allowances have been made, both by us and the DES, for both high and low social class variables.

Criticism of Marks and Pomian-Srzednicki

Although several aspects were taken into account in the second study in response to comments made by others, for example the social class composition and the representativeness of schools, there are still some aspects of the studies that can be questioned. In both studies, a selectivity index was devised by using the ratio of the number of pupils entering grammar schools against the size of the whole age cohort entering secondary schools in that LA. This was for 1976 in the first study and 1977 in the second study when these pupils entered secondary school. These percentages may have changed during the time these pupils were in secondary school. It is possible that some schools could have changed their status and some pupils would have changed schools within the five years. We also note, as in the previous paper, having defined a selectivity index, the authors then go on to ignore it in their attempt to create wholly selective and wholly comprehensive systems for comparison. Their principal conclusion is again based on the assumptions of 25% and 30% selectivity and thus should be interpreted with caution as in the previous study. We have already noted in Chapter 1, how much variation there is in the level of selection across the country in 2005; we do not know the detail of how much variation there was in 1976-1982. More so, the use of the LA as the unit of analysis is also problematic, though it is possible that the number of students crossing LA borders in 1976-1982 was low enough for this not to be such an issue as it is today (see Chapter 6).

We also note that the results show the selective system in a better light for the OC1 passes than the points per pupil. The OC1 passes only include the more able pupils at age 16, and it may well be that the selective system was doing well for these pupils, but not so well for the cohort of pupils viewed as a whole.

3.2. *More recent studies*

We know of no empirical studies with new data carried out over the fifteen years between 1985 and 2000. As outlined in Chapter 1, the debate on the most effective way to organise secondary education continued throughout this time with various political stances being seen in the introduction of new legislation.

Kerckhoff *et al* (1996) did revisit the NCDS data sweep of 1974 and carried out their own analysis (reported in Crook *et al*, 1999) in which, after controlling for socio-economic background and prior achievement, they found no significant differences in the average achievement of pupils in selective and comprehensive schools. However, they did report clear evidence that high ability pupils performed at higher levels in selective systems and low ability pupils performed at higher levels in comprehensive schools, although for most students the differences were said to be small.

Writing in 1999, Crook *et al* noted that the question *which type of school, selective or comprehensive*, is more effective had failed to produce a consensus from research carried out over the previous 30 years. They noted that some of the methodological difficulties are fundamental, such as the definition of a comprehensive school, and whether or not such a school is creamed. They quote Walford as stating in 1994 that the number of selective schools has declined to a level where further meaningful comparisons between systems are impossible.

However, research continued with three major studies reported in 2000 (Jesson), 2003 (Schagen and Schagen) and 2004 (Atkinson *et al*). Before reviewing these studies we illustrate that the political debate was still very much alive at the turn of the century with a parliamentary question and its answer.

3.2.1. House of Lords, 2000

On 15 March 2000, Baroness Blatch (Conservative) asked:

> Whether Her Majesty's Government will now support the continuance of the principle of selection by ability in some maintained British schools?

Baroness Blackstone replied on behalf of the Labour Government

> The government believe that a non-selective system is one that is likely to serve the needs of all children. I do not believe there is any support for a move towards a selective system….Most parents do not wish to have the opportunity to send their children to schools which select by ability….It is our policy to oppose selection on ability and through interviews.

Baroness Blackstone was asked

> to confirm that the average performance of the top quarter of pupils at comprehensive schools is just as good as that achieved in grammar schools, and was proof that comprehensives are at least as effective as grammar schools in providing a first class education.

She responded,

> the top 24% of pupils in maintained comprehensive schools is slightly higher than that in grammar schools. The percentage of

> pupils in grammar schools who achieve five-plus A to C at GCSE is 95.4%. The figure for a similar level of achievement in comprehensive schools is 100%. That demonstrates without a doubt that comprehensive schools are doing very well.

Later Baroness Blatch asked for the statistics that justified the statement.

In the House of Lords, 6 April 2000, Lord Bach gave the answer:

> The statement reflects the comparative performance of two groups of pupils. The first group included all 15 year-olds in grammar schools. The second group included the top 25% of 15 year-old pupils in comprehensive schools—based on their total GCSE/GNVQ point score (in order not to include and exclude some pupils with the same point score, the actual percentage of pupils included was 24.5%).

The results were tabulated

Performance	grammar	comprehensive
Percentage achieving 5+ grades A*-C	96.4%	100%
Percentage achieving 1+ grades A*-C	99.6%	100%
Average point score per 15 year-old pupil	60.7	60.9

Lord Bach stated,

> this analysis was carried out by the department's (DfEE) Analytical Services. It is based on the data collected for the secondary school performance tables published in 1999. The analysis was carried out from a pupil level file and based on the results of all 15 year-old pupils in grammar schools and the top 25% of pupils in comprehensive schools. This percentage was chosen because it reflects the proportion of pupils typically taken in by grammar schools in wholly selective LAs. The pupils in the comprehensive schools were ranked according to their total GCSE/GNVQ point scores.

Although the statistics provided by the DfEE appear to show the benefit of the comprehensive system, the method by which they tried to identify comparable populations in selective and comprehensive systems is so flawed as to be wholly without credibility. There are two main reasons for this.

Firstly, the results would be very sensitive to the percentage set for the proportion of pupils in comprehensive schools deemed to be equivalent to those in grammar schools. Setting the proportion to be 30% would significantly reduce the apparent performance of comprehensive schools, while choosing the top 20% as representative would make them look even better. Yet the choice of 25% (or more accurately 24.5%) is totally arbitrary, as grammar schools in different parts of the country take different proportions of the local population, and local populations vary appreciably anyway in their ability distributions.

Secondly, the top 25% at age 11 will not be the top 25% at age 15. Grammar schools select pupils on the basis of their ability at age 11 but other factors will influence the attainment of these pupils at age 15. The extent to which these two populations can differ has already been illustrated in Chapter 2. Comparing the top performers in comprehensive schools at age at 15 with the whole grammar school population at the same age will inevitably favour the former.

In addition, we note that the data are limited to this top 25%, so the claim that *"That demonstrates without a doubt that comprehensive schools are doing very well"* is totally overstated.

3.2.2. *Jesson (2000): The comparative evaluation of GCSE value added performance by type of school and LA*[1]

In this paper Jesson reports on analyses using national data sets that had become available for the first time (1998) to make comparisons between the GCSE results of LAs which were fully comprehensive, with those of LAs which still retained some grammar and secondary modern schools. He made use of analyses based on value added techniques to challenge the claim made by some that selective LAs get the best results. He found no evidence for the superiority of either grammar schools themselves, nor selective systems of educational provision, and suggested his results showed that comprehensive systems produced better GCSE performance.

Jesson's Conclusions

- The evidence 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 that similar pupils in fully comprehensive systems.

Jesson's paper essentially reports on four separate, but related, studies.

We describe, and criticise each one in turn.

Study 1: Using data published in the DfEE Value Added Pilot Study: 1998

Value added was calculated as a measure of school effect from a child's attainment at key stage 3 in 1996, through to his/her GCSE results in 1998.

186 state schools volunteered to take part in this pilot study; comprising

comprehensive schools	157
grammar schools	19
modern schools	10

(DfEE used the term modern rather than secondary modern)

These schools were said to be broadly representative.

Pupils were classified on the basis of their average key stage 3 results

low prior attainment	level 4 or below
average prior attainment	between level 4 and 5.66
high prior attainment	above level 5.66

[1] Discussion papers in economics.
Centre for Performance and Resource Management; University of York

Jesson argued that it was appropriate to use the high prior attainment group to compare grammar schools with other types of school, as grammar schools select the most able pupils.

He noted that each of the ten modern schools had sufficient pupils in this category to include them in the study.

Table 9: Percentage of schools, split by school type and progress measure.

Progress Measure	Comprehensive Schools	Grammar Schools	Grammar and 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%

Table 9 is reproduced from the report, and shows the percentage of schools ascribed to each of 5 groups, A, B, C, D and E

The progress measure is between attainment at Key Stage 3 and GCSE. (NB: there is no detail of how this was calculated, nor how the five groups were defined).

Jesson argued that the combined grammar and modern schools, related, in the main, to performance in 'selective' areas. (his quotation marks) the other two columns providing a direct comparison of comprehensive and grammar schools.

Jesson noted that pupils in these comprehensive schools appeared to make at least as good progress as their counterparts in grammar schools. In particular 10% of comprehensive schools were rated A for their performance, compared to 0% for the grammar schools. He also noted that for categories A and B combined, 1 in 3 comprehensive schools were in this category, compared to only 1 in 7 for selective systems. Jesson acknowledged the small sample, but argues that this analysis of the most able pupils represented a challenge to the claims that grammar schools provide better performance.

We note that although these volunteer schools are described as *broadly representative*, we have no information about where in England they come from. Combining the grammar schools and modern schools and calling it a *selective area*, or even a *selective system*, is thus questionable.

There is no detail of how the value added was calculated, but the definition of a most able pupils as Key Stage 3 being greater than 5.7, is itself very broad. The 'median line' approach that the DfEE subsequently adopted has been shown to underestimate the progress generally made in schools with high ability intakes (Critchlow and Coe, 1999). We do not know how many children were in each category. Moreover, Key Stage 3 to GCSE represents only two years of schooling, and takes no account of progress from entering the school in year 7 to taking the tests in year 9.

Study 2: Using data published in the 1998 GCSE/GNVQ performance tables.

Jesson notes this was the first ever, national value added evaluation of secondary school performance. He made use of the analysis made by the DfEE. He quotes from the DfEE standards internet site:

> Schools were placed in one of 27 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 measured.

We find this difficult to follow but understand it to mean:

Define 27 bands across the 1996 KS3 performance (there is no detail of how the bands are defined)

Allocate each school to a band by its average KS3 performance

Calculate the median of each band.

Calculate the difference between the median and each school's score.

Rank all schools according to this difference

Classify schools by 95th, 75th, 25th, 5th percentile in this ranking

Jesson continued: the performance tables contained a ✓ for each school adjudged to be in the top 25% of schools serving 'similar' pupils (his quote marks).

We note we are not told the criteria for 'similar'; is it in the same band as defined by the percentiles or the same one of the 27 bands defined to classify Key Stage 3 performance, …or something else? This matters, because Jesson now goes onto discuss LAs where grammar schools would be subject to parental ballots on their future status, and the number of ✓s awarded to schools in these areas. These are shown in the table, Table 10, reproduced below.

To Jesson's original Table (i.e. DfEE 1998), we have added 2003 figures (shown in bold), for the total number of schools, and included the number of modern schools and comprehensive schools in these LAs as defined in the DfEE 2003 performance tables. Jesson says his table shows how well these grammar schools and their associated secondary moderns faired. In looking at the 2003 figures this is ambiguous as there are 67 comprehensive schools. It could be argued that because these 67 schools are in these selective LAs are they really secondary modern schools, but they are designated comprehensive by the DfEE.

Jesson made two observations from the table:

1. Overall, the last two lines of the table show that selective areas had fewer good performing schools than areas which had other forms (mainly comprehensive) of organisation.

He notes 16% of schools in selective areas 'did well', compared to over 27% in mainly comprehensive areas (we note the use of the term *mainly* presumably as the other 70 grammar schools are included in these comprehensive areas).

2. Taking grammar schools on their own, only 13 out of the 94 in these ten LAs 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 a performance as other types of schools.

Table 10: Number of different school types split by LA

LA	Total number of schools		Number of grammar schools	**Number of modern & comps**		Number of schools with a ✓ 1998	Percentage with a ✓	Number of grammar ✓	Percentage with ✓
	1998	**2003**							
Bexley	16	**16**	4	**5**	**7**	7	44%	0	0%
Bucks	36	**34**	13	**21**	**0**	8	22%	4	31%
Kent	105	**102**	33	**47**	**22**	14	13%	5	15%
Lincs	64	**63**	15	**32**	**16**	8	13%	1	7%
Medway	20	**19**	6	**10**	**3**	3	15%	0	0%
Slough	11	**11**	4	**6**	**1**	4	36%	2	50%
Southend	14	**12**	4	**4**	**4**	1	7%	0	0%
Sutton	14	**14**	5	**0**	**9**	3	21%	1	20%
Torbay	8	**8**	3	**0**	**5**	0	0%	0	0%
Trafford	18	**18**	7	**11**	**0**	0	0%	0	0%
All select areas	306	**297**	94	**136**	**67**	48	16%	13	14%
All other	approx 2800		70			750	27%		

Jesson concluded: these indications offer a further challenge to the validity of claims that grammar schools provided superior performance for pupils overall when compared to comprehensive schools, and from Table 1 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.

We note that in this analysis the concept of 'good performing' or 'doing better' is based on the value added from Key Stage 3 to GCSE. Again we do not know how this was calculated, but we do know it was the DfEE *pilot* study. We also know that many children in grammar schools score much higher than level 5.66 at KS3, some reaching levels 7 and 8. Such children have little or no scope to add value at GCSE. The GCSE examinations, even at the A* grade level, prevents them from demonstrating progress as measured here. The GCSE provides a ceiling effect for these pupils. What may be happening is that progress between years 7 and 9 is far less in comprehensive schools, so that in years 10 and 11 there is far more scope for children to add value, and thus according to the criteria used here, to out perform their grammar school peers. To use these value added measures in this way is clearly flawed.

Prais (2001) noted: His (Jesson's) sums are far from implying that grammar schools are, on the whole, less successful in true educational terms than comprehensive schools; as noted; they may imply no more than that

grammar schools start from a higher initial position at age 14 and are less variable in their GCSE results.

We also note how difficult it is to define what is meant by a selective area or selective system, yet alone a fully selective system.

Study 3: Using national data on prior attainment and GCSE outcomes

Jesson notes that the DfEE national data sets allowed him to match 500,000 children who attended state schools from Key Stage 3 in 1996 to their individual GCSE results in 1998.

He identified four types of school within the data with number of schools and pupils as shown in Table 11

Table 11: School types and number of pupils

School type	Number of pupils	Percentage	Number of schools	percentage
Comprehensive	448725	89.4%	2714	87.1%
Secondary modern (selective LAs)	34370	6.9%	246	7.9%
'isolated' grammar schools	5696	1.1%	54	1.7%
Grammar (selective LAs)	12895	2.6%	101	3.2%

Jesson noted: secondary modern schools included all non-grammar schools in those areas where selection of around 20% or more of the pupil population was in place.

We note he doesn't say how these areas are defined; presumably it is LAs.

He cites the three grammar schools in North Yorkshire to explain what he means by isolated grammar schools.

We note this would imply that presumably any grammar schools that are not within his selective areas are isolated. If so, it would mean, the Birmingham grammar schools are isolated.

Pupils were classified in terms of prior attainment on the basis of their average Key Stage 3 results; i.e. each child is deemed to be at a level between 3 and 8 inclusive for each of mathematics, English and science (he included children who had been attributed a level for at least two of the three subjects). The bands were chosen to get groups of similar sizes.

low	level 3.67 or below
below average	between levels 3.67 and 4.5
average	between levels 4.5 and 5.2
above average	between levels 5.2 and 5.67
high	level 5.67 and above

He tabulated his results. Table 12 shows the distribution of pupils across prior attainment groups in each type of school.

Table 12: Distribution of pupils across prior attainment groups in each type of school

	high	> ave	ave	< ave	low	School key stage 3	
School type						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 LAs)	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

Jesson focuses his discussion on the standard deviations; i.e. a measure of the degree of variability about the mean score, noting it is smaller for grammar schools in his selective LAs compared to isolated grammar schools. He suggests this shows that local conditions of entry into such schools may be much more variable than those applying to grammar schools in selective LAs. He describes the areas in which isolated grammar schools are situated as predominantly comprehensive. Jesson gives the combined mean of secondary modern schools and grammar schools in selective areas as 5.04 but makes no further comment on the mean values.

We note 5.04 is greater than 4.74 given for comprehensive schools, and the mean for both Jesson's two types of grammar school is over 1.5 Key Stage 3 levels above the comprehensive schools. He also notes that the data shown in Table 11 and Table 12 do not represent a full or complete account as they are based on information relating to the final two years of secondary education, nor do they take socio-economic conditions in to account (this is somewhat confused as the data refer to Key Stage 3; he is yet to discuss value-added at GCSE). It is also somewhat odd to use a standard deviation found from these data, obtained from pupils in year 9, to discuss entry to the schools some 3 years earlier.

In continuing this study Jesson contrasts what he calls non-selective and selective systems of educational provision the latter being a combination of grammar and secondary modern schools. He states this allows contrasts to be drawn directly between outcomes from the two types of system; 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.

Jesson defined a selective area as an LA in which around 20% or more of its year 11 cohort of pupils were in grammar schools.

Table 13 below shows Jesson's figure for 1998 based on year 11 pupils, to which we have added figures for 2003, with percentages based on year 9. We have also added the number of different types of school (according to DfEE designation) in Jesson's selective areas. The percentages in grammar schools (year 11 in 1998; year 9 in 2003) are very similar with the exception of Trafford.

Table 13: Descriptive statistics of schools in 'selective' LAs

LA	Grammar School	Secondary Modern	Average KS3	**Grammar school**	**Secondary modern**	**Comprehensive**	
	1998			**2003**			Percent in grammar
	Percentage of pupils			**Number of schools**			
Bexley	22%	78%	4.93	**4**	**5**	**7**	23%
Sutton	31%	69%	5.47	**5**	**0**	**9**	29%
Wirral	22%	78%	4.92	**6**	**5**	**11**	22%
Trafford	30%	70%	5.06	**7**	**11**	**0**	42%
Bucks	40%	60%	5.29	**13**	**21**	**0**	38%
Poole	20%	80%	5.26	**2**	**3**	**3**	20%
Bournemouth	19%	81%	5.12	**2**	**7**	**1**	17%
Reading	23%	77%	5.04	**2**	**0**	**5**	21%
Slough	36%	64%	4.82	**4**	**6**	**1**	38%
Torbay	29%	71%	5.33	**3**	**0**	**5**	26%
Southend	30%	70%	5.19	**4**	**4**	**4**	29%
Kent	29%	71%	5.05	**33**	**47**	**22**	28%
Medway	25%	75%	4.92	**6**	**10**	**3**	26%
Lincs	20%	80%	4.99	**15**	**32**	**16**	23%
All LAs	3.7%	6.9%	4.80				
				106	**151**	**87**	

We note that the figures for 2003 question Jesson's statement above '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'; 87 of the schools in this area are designated comprehensive (2003) and 100 of the 151 secondary modern schools are in three LAs (Kent, Buckinghamshire, Lincolnshire). It would be more accurate to say, 'not educated in grammar schools'.

Jesson notes that the average KS3 level for all of his selective areas is above 4.80, the figure he gives for all LAs in England and concludes this places even greater emphasis to the need to evaluate school and pupil performance by Value Added means.

He then presents the following table, Table 14, relating pupils' average GCSE performance (1998) to their prior attainment at Key Stage 3 (1996)

Table 14: Pupils' average GCSE performance (1998) related to attainment at Key Stage 3 (1996)

Prior attainment group	Comprehensive	Selective LAs	Comprehensive	Selective LAs
	GCSE Point 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 s.d. is the standard deviation, a measure of the variability about the average values)

Jesson notes the average performance of comprehensive school pupils was higher than that for selective system pupils for all but the most able pupils.

We note Jesson does not say how the GCSE point score has been calculated, but in 1998 it would be based on 8 points for an A*, 7 for an A to 1 point for a

grade G. On that basis the differences are of the order of 1 or 2 points, implying, according to these figures, that the two systems are equitable in performance. This is not surprising given our criticism of how Jesson's defined his selective systems.

Jesson claims this is essentially a value added evaluation of performance, in that it is related to prior attainment, although no value added calculations were actually carried out.

Study 4: Outcomes for pupils in Comprehensive and Selective Schools

In his fourth study Jesson used multi level regression analysis to predict the GCSE points of a male or female pupil from his or her Key Stage 3 average level. He claimed these two factors of gender and prior attainment, accounted for almost two thirds of the variability in pupils GCSE outcomes. He quoted R^2=62.4%.

Note, R^2 is a statistic involved in regression analysis that gives a measure of confidence in the results. Many factors other than gender and prior attainment will affect a particular pupil's actual GCSE results, and the stated result of accounting for nearly two thirds of the variability with just two factors does seem remarkably high.

Jesson gave his predicted GCSE points formula for a female pupil, as

$$37.1 + 14.4 \text{ (average KS3 score} - 4.8)$$

with male pupils being 3 points less.

Jesson stated that 1.1 points should be deducted for pupils in selective systems and he concluded from this that pupils in selective systems of school organization performed slightly less well than their peers in comprehensive systems.

We note the difference of 1.1 points is equivalent to about a grade G. Jesson's results again show his comprehensive and selective systems performing at about the same level.

Jesson also stated he had carried out a similar analysis on pupils' performance on 5 or more GCSE grade A* to C passes. He gives no details but states pupils in selective systems achieved around two percentage points lower than their peers in comprehensive schools.

We note this would appear to be consistent with Table 15, but there no allowance is made for what subjects these 5 passes may be in.

Jesson continued to describe a value added measure he used to compare the performance of schools within his comprehensive and selective system. At the heart of this is a comparison of each pupil's actual performance at GCSE compared to what he/she would have achieved had he/she performed in line with the national average for pupils like them. These differences were then summed for each school to give a value added measure as to whether the school did better or worse than expected.

He presented his results in the following table, using five categories A, B, C, D and E defined as:

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

Table 15: Breakdown of value added results, split by LA organisation

LA organisation	A	B	C	D	E
Comprehensive	4.8%	8.4%	76.1%	8.0%	2.7%
Selective	0.9%	5.2%	77.5%	11.0%	5.5%

Apart from saying they are based on normal distribution z scores as above, the details are not given; he states these bands were chosen to give a reasonably balanced description of schools' performance using conventional statistical practice for the definition of category C.

He notes some 13% of comprehensive schools are in categories A and B, over twice that for selective system schools. He said this indicated a clear advantage for schools in LAs where the organization is comprehensive in nature.

We note there is a lot of detail missing in the above. The value added is based on *a national average for pupils like them.* We do not know how *pupils like them* are defined. Is this Key Stage 3 prior attainment as in the 5 bands of his earlier tabulated results, or in 27 bands of the DfEE, or some other definition. Summing differences over a whole school can only give a crude indication of the performance which would be expected to vary considerably between individual pupils. Also, as noted earlier, pupils in grammar schools have already made considerable progress by the time they reach Key Stage 3, so may add little or no value between then and taking GCSE.

Jesson completes this study by finding what percentage of the pupils from each of the types of school he identified in Table 12 fall into each of his categories A, B, C, D and E for value added.

The results are shown in Table 16.

We note that from Table 11, that 89.4% of the pupils of these pupils are in Jesson's comprehensive schools category.

Table 16: Percentage of pupils, split by value added category and school type

	A	B	C	D	E
Comprehensive	5%	8%	75%	8%	3%
Secondary modern	1%	4%	76%	12%	7%
Isolated grammar	0%	15%	81%	2%	2%
All grammars	0%	10%	82%	6%	2%
Selective systems	1%	5%	77%	11%	6%

Jesson highlights that in category A comprehensive schools outperformed all other types of school and comprehensive schools, with 13% in categories A and B outperformed the 10% of grammar schools in these categories.

Jesson concluded:

> 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.

And 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. He noted in particular that secondary modern schools are an essential part of selective systems.

Jesson states his results raise a substantial challenge to the claim that grammar schools produce better examination results. In almost all of the comparisons made above the evidence points in the opposite direction.

In answering this challenge we note: there is an inherent difficulty in defining a selective system or selective area when distribution of grammar schools around the country is not uniform, nor is the percentage of pupils from any one LA who attend them.

The very disproportionate sample sizes used in comparing comprehensive schools taken from all over England to Jesson's selective areas.

The ceiling effect in value added between Key Stage 3 and GCSE, which means some pupils with outstanding results at GCSE cannot demonstrate value added, because they also had outstanding results at Key Stage 3.

The value added measures reported by Jesson are a comparison of schools and not individual pupils, based on an aggregating process that is only a crude measure. It is also limited to two years of schooling, and takes only formal assessment as its starting and finishing point. Other aspects of school, such as development of attitude and motivation towards education and subjects for post-16 study, are not considered.

Jesson acknowledges that his study is only for the years 1996-1998, and that further work needs to be done, particularly over five years from Key Stage 2 to GCSE.

The value of his paper is in showing the potential that the national data sets have for this type of investigation. His conclusions as such are overstated and his results should be stated as tentative and in need of further investigation as more data becomes available, and statistical techniques to analyse them are developed.

Finally we quote from Prais (2001) on this value added study:

> The measuring rods (value added) are thoroughly unsatisfactory because:
>
> 1) pupils in grammar schools on average start at 14 from a higher National Curriculum level and attain higher and more uniform GCSE grades at age 16.
>
> 2) there is a ceiling effect which prevents the true extent of high attaining pupils performance becoming evident
>
> 3) we have to be careful as to how greater variability within comprehensive schools affects the outcome of such calculations.

Prais concluded any further calculations on value added by the DfEE will need to be carried out more tentatively, will have to be set out more explicitly, and be interpreted more cautiously. As Jesson used the DfEE data, these comments also apply to his results.

3.2.3. *Jesson (2001): Selective systems of education – blueprint for lower standards?*

Jesson followed his 2000 paper with a summative version, published in the above journal. In this he revisits the 1996-1998 DfEE pilot study on value added, comparing selective, partially selective and non-selective areas of the country. He again emphases that comparison of educational performance across different forms of school organisation clearly need to take account of the achievements of ALL pupils (*his emphasis*) again arguing that those in favour of grammar schools do not consider the results of pupils in the associated secondary modern schools. Jesson claims that selective school systems have failed the nation and that grammar schools and secondary modern schools have no place in this new world.

Jesson's conclusions

- We have found clear evidence that selective school systems depress educational performance.
- We need to avoid the unnecessary division of pupils and their schools into selective and non-selective types, because this only exacerbates lower performance.

Jesson describes the present situation. He notes the non uniform distribution of the grammar schools in England saying 108 are located in 15 LAs, with over half of these in Kent, Lincolnshire and Buckinghamshire. Most areas,116 LAs do not organise secondary schools on a selective basis, instead schools in these area serve, to a greater or lesser extent, the local communities in which they are situated and parents are free to make choices amongst <u>all</u> (*his emphasis*) of the available schools.

He summarised his data in Table 17

Table 17: Distribution of schools in 'selective', 'partially-selective' and 'non-selective' areas

System	LAs	No of schools	Grammar	Non-selective	Percent in grammar
Selective	15	350	108	242	28%
Partially selective	19	637	54	583	6%
Non-selective	116	2081	0	2081	0%

We note that according to the DfES data for 2003 Jesson's figures are not correct. (see Table 1, p12). There are 164 grammar schools. There are 15 LAs which have 4 or more grammar schools, totalling 125. Of these 33 are in Kent, 15 in Lincolnshire and 13 in Buckinghamshire, totalling 61. The number of LAs with no grammar school is 114. Jesson's choice of words implies the LAs where the grammar schools are situated have made a conscious decision to organise their schools along selective lines. This would seem to ignore the recent history of local school organisation from about 1980, and the shift of power from LAs to school governing bodies and their powers to make decisions on entry policy. Jesson also suggests not all parents can apply for their children to attend a grammar school; of course all parents can apply, but not all applicants will be successful.

Jesson has not defined what he means by selective, partially selective and non selective areas. In his previous paper he defined selective area as about 20% or more pupils in an LA attending a grammar school.

We have reproduced his Table 1 as Table 18 here, using the 2003 DfEE data. This is based on two definitions of selective: 'a' an LA is selective if it has no comprehensive schools or 'b' following Jesson, if it has about 20% or more pupils in a grammar school (this is taken as 17% or more from Table 1, Chapter 1, and this produces figures close to Jesson's). We have not attempted to calculate percent in grammar, because as we suggested in Chapter 1 this is a meaningless figure, a view we reinforce with strong evidence in Chapter 6) The final two columns in Table 18 below show the split between secondary modern schools and comprehensive schools, in Jesson's non selective schools (following his definition of selective).

Table 18: A comparison of definitions of 'selectivity'

System	LAs		No of schools		Grammar		Non-selective		Sec mod	comp
	a	b	a	b	a	b	a	b		
Selective	3	15	62	354	22	108	40	246	159	87
Partially selective	33	21	999	707	142	56	857	651	21	630
Non-selective	114	114	2067	2067	0	0	2067	2067	1	2066

The major criticism of Jesson here is that the figures in his Table 1 are at best misleading, and could be said to be totally wrong.

Jesson says that those pupils not selected for grammar school, go to non-selective schools, often, but not exclusively called secondary modern schools. He states: these latter educate only those who have failed to gain entry to grammar school. Selection in the fifteen selective LAs therefore embraces two types of school, **grammar** schools and **secondary modern** schools. *His emphasis*.

We note from Table 1 (p12) that there are 87 comprehensive schools in Jesson's selective areas. He can call these schools secondary moderns, but he is clearly overstating the case about his selective LAs embracing two types of school. He also assumes that every parent/child in these areas wants to go to a grammar school; some might opt for a comprehensive school by choice and thus would not have failed to gain entry to a grammar school.

Jesson makes the statement: perhaps one reason for the poor performance of the selective system as a whole derives from the impact on these pupils of the educational progress of being classified as failures at age 11. This feature is of course completely absent in non-selective areas.

We note that Jesson puts forward no supporting evidence here for his statement about poor performance, and he is presumptuous in saying those pupils who do not gain entry to grammar schools are classified as failures. Again there is nothing to support this. His statement that this is completely absent in non-selective areas is fallacious as CTCs and Specialist Schools can both have an element of selection on their entry policies. He seems to assume that elsewhere in the country any child can gain entry to any state school of his/her parents' wishes.

Jesson continues to compare Key Stage 3 average levels and GCSE results in his selective areas, now calling them fully selective areas. He uses the same data as in his 2000 paper, but he chose to present it differently.

We note that Jesson's Table 19 should be compared with Table 11 taken from his 2000 paper.

Table 19: The 'ability' profile of LAs having each type of secondary school organisation

	LAs	Below average	Average	Above average	Mean
					KS3 points
Selective	15	28%	19%	53%	34.5
Partially selective	19	32%	21%	48%	33.5
Non-selective	116	36%	22%	42%	32.5
Overall	150	35%	21%	44%	32.9

Jesson has combined his former five Key Stage 3 prior attainment groups into three, and presented the Key Stage 3 attainment using points rather than levels.

It is noticeable that when using points the school systems look very similar, and do not show the clear difference in favour of grammar schools seen in the previous table.

Jesson now gives a brief discussion of the advantages of value added but makes a very summative table of his results from his fourth study in the 2000 paper.

Table 20: The performance of average pupils in LAs of differing types of organisation

	Percent 5 or more A* to C	GCSE points
Non selective 116 LAs	45.3%	37.0
Selective 15 LAs	43.1%	35.9
Non selective advantage	+2.2%	+1.1

Jesson concludes: In selective areas pupils who are not selected go into secondary modern schools and their GCSE performance actually depresses the communities overall performance. That is **selective systems result in lower educational performance** (his emphasis).

All our criticisms of Jesson's 2000 paper continue to apply here, but in particular we reiterate that these differences are very small, we do not know from which GCSE subjects they were calculated, the sample size is very unbalanced and there is a great inherent difficulty in defining a selective system. So although his conclusion is worthy of further investigation, it has in no way been demonstrated to be true.

Jesson remarks in conclusion: we need also to build on the example set by many of the newly designated (non-selective) specialist schools which have shown significant and substantial improvements in performance along with highly significant value added assessments of their performance.

It is interesting that he describes these schools as non-selective when they have supposedly been set up for children who demonstrate certain aptitudes, and can select up to 10% of their entry, as discussed in Chapter 1.

3.2.4. *Jesson (2007): A ladder of opportunity? The pupil intake and performance of England's grammar schools*

In a recent review, Jesson discusses and challenges 'recent debates, particularly within the Conservative party' of whether children from poor family backgrounds are offered a 'ladder of opportunity' within grammar schools to achieve higher levels of success than would be expected compared to their state educated peers. He also discusses whether the good exam outcomes of grammar schools are sufficient for the argument towards their retention or expansion.

Discussing the socio-economic context of grammar schools, he uses the proportion of pupils eligible for free school meals in grammar and other schools in each different type of local authority (no grammar schools, partially selective and fully selective) to show that very few such pupils are educated in grammar schools (2%) while figures in non-selective schools are close to the national average (13% in 'no grammar school' authorities, 14% in partially selective and 12% in fully selective ones). He concludes from this that the polarisation between grammar and other schools is considerable and as grammar schools are 'in no way having a representative sample of the pupils in their areas, … this is hardly providing a ladder of opportunity for disadvantaged pupils'. Jesson then uses the IDACI (the 'income deprivation affecting children index') to identify the extent to which schools recruit pupils from 'severely disadvantaged' (the highest 20%) and 'low disadvantaged' (the lowest 20%) communities and again states that grammar school intakes are clearly polarised: 'over one third of their pupils come from low disadvantage communities; and only 5% from those with severe disadvantage, compared with national figures of 20% in each category'. He again concludes that grammar schools do not provide less advantaged pupils with a 'ladder of opportunity'.

Jesson then argues that grammar school selection processes demonstrate another bias against children from disadvantaged backgrounds as more affluent parents who would be able to send their children to fee-paying schools are in a more favourable position in securing places at local grammar schools over similarly able pupils from state schools. He presents a table to demonstrate this bias:

Table 21: Entry to grammar schools from non-state schools: Extracted from Jesson (2007)

Local authority type	Percentage of pupils from state and non-state schools			
No grammar schools (114)	State junior		Non-state	
	95%		5%	
	Grammar schools		Other schools	
	State	Non-state	State	Non-state
Partially selective (21)	85%	15%	96%	4%
Fully selective (15)	87%	13%	95%	5%
Total selective (36)	86%	14%	96%	4%

Jesson argues that non-state schools have the opportunity to coach their pupils for grammar school selection tests in the same year where pupils in

state schools take the key stage 2 tests. Coaching pupils in state schools for these tests would have to take place outside the national curriculum time and would require the use of private tutors which disadvantaged families would not be able to afford. He again concludes that this cycle of deprivation and of limited opportunities is perpetuated as grammar schools are far from providing the ladder of opportunity for children from disadvantaged backgrounds that many of their supporters claim. Indeed, a more appropriate description might be that they are 'ghettos of the advantaged', doing little to alleviate the divisions in the society they are intended to serve (Jesson, 2007).

Finally, Jesson goes on to discuss how the official classification of schools by a 'value-added' measure which compares schools on their pupils' progress rather than on their 'absolute' performance, has shed new light on the apparent high performing outcomes of grammar schools. It shows that on the whole their performance is as good as would be expected given their pupils' ability on entry and it is no better than the overall performance of non-selective comprehensive schools. He also notes that there are 16 grammar schools which have been assessed as providing below average progress for their pupils and he states that these schools have been found to have higher proportions of pupils eligible for free school meals and from severely disadvantaged communities ones. He intends to make the point here that, given their intake, grammar schools are no better than state schools in the country. Jesson notes, however, that this finding needs to be interpreted with extreme caution due to the relatively small sample of pupils involved. He concludes that only a very small number of disadvantaged pupils are offered a ladder of opportunity through grammar schools as these schools tend to favour pupils from more prosperous communities.

It is difficult to draw any firm conclusions from Jesson's discussion paper as no information is available on the sampling procedure, on pupil, school or LA selection or on the actual method of analysis that was followed. However, it has been mentioned here as an example of the recent debate on the under-representation of children from disadvantaged backgrounds in the better schools (see also BBC, 2006; 2007a) often leading to the failure to improve social (BBC, 2007b) and intergenerational mobility (Blanden *et al*, 2005).

3.2.5. Prais (2001): Grammar School's Achievements and the DfEE's Measure of Value-added: an attempt at clarification

Prais is critical of the methodology of the DfEE pilot on value added 1996-1998, and in particular the conclusions that Jesson draws from this data in his 2000 paper. Prais makes his own interpretation of the value added analysis and concludes the data show grammar schools out performing comprehensive schools for high attaining pupils.

We have already quoted from Prais in our criticism of Jesson's analysis and conclusions from the DfEE pilot study. Prais makes another point here, that attainment at Key Stage 3 is based on only the three subjects of English, mathematics and science, whereas the curriculum for high attaining pupils will contain much more than this, particularly in a grammar school, where for example, some may take a second foreign language. A pupil's Key Stage 3 score is thus only a limited measure of his/her attainment at age 14 yet this formed the baseline for measuring value added at age 16.

Prais says the 25 grammar schools (Jesson said there were 19) and other selective schools in the DfEE sample (he included independent schools) had an

average of 6.2 per pupil, or 1.5 levels above that for comprehensive schools. As one level is notionally equivalent to two years of schooling, Prais concluded this evidence shows grammar school pupils to be about three years ahead of comprehensive school pupils at Key Stage 3. He notes at age 16, when comparing GCSE point score, the point score for grammar schools, 57.5, is 24 points greater than for comprehensive schools. He comments this is not surprising given the difference in pupils attending the two types of school.

We note that this underlines how difficult it is for able children to demonstrate value added starting from the DfEE Key Stage 3 measure.

Whereas Jesson had compared only pupils who had high prior attainment, Prais compared all pupils, and obtained a value added score of –1.4 for comprehensives and –0.7 for grammar schools. So when comparing all pupils Prais concluded, grammar schools do better but he again pointed out the severe limitations of the way value added was measured, in particular pointing out that pupils who achieved 8 grade A*s at GCSE could have negative value added indicating a major flaw in this measure.

He describes the methodology used by the DfEE and followed by Jesson, of using the ticks, as precarious. However, that said he produces figures for the top ability group of –1.7 for comprehensive schools and –0.6 for grammar schools, implying that by this methodology grammar schools add more value.

Prais noted another methodological problem with value added in the assumption of a linear scale, and that we don't know if improving from grade F to G is the same as improving from grade A to A* but the value added measure assumes it is.

Prais also noted confusion and uncertainty in how the prior attainment groups had been defined both in schools and nationally. Prais called them technical difficulties, but concluded the DfEE pilot study left us very little forward.

We note that Prais really points out the difficulties inherent in the DfEEs pilot study, and so no significance should be read into the results. This reinforces the shaky ground on which Jesson draws his conclusions in his two papers.

3.2.6. *Yang and Woodhouse (2001): Progress from GCSE to A and AS Level: institutional and gender differences, and trends over time.*

Although we would not describe this as a major study into the effectiveness of selective systems, we have included it as it is the only study we found that considers post 16 performance.

Yang and Woodhouse (2001) investigated the relationship between students' results obtained in examinations for the General Certificate of Education (GCE) at Advanced and Advanced Supplementary (A/AS) level and those obtained by the same students two years earlier in GCSE examinations. The main aim of the study was to describe educational progress, and how this differed for different kinds of students and in different establishments.

Data

The results of four cohorts of students who sat AS/A level examinations between, 1994 and 1997 and who were 18 during the year. The students

attended a recognised educational establishment in England, and entered at least one examination at A/AS level examinations. The data actually used comprised 696,660 students attending 2,794 establishments between them. The establishments included LA maintained selective and non selective schools, independent schools, sixth form colleges and further education colleges. Selective areas were not considered.

Explanatory variables

Gender, age, establishment type, region and time, and measures of performance at GCSE. Age was measured relative to 18 years and 6 months, and time refers to when the examinations were sat. GCSE performance was measured by using a combination of mean GCSE score and the number of GCSEs taken. The GCSE score was based on 7 points for grade A* and A, 6 for grade B down to 1 for grade G.

Outcome variable

This was a composite points score based on individual subject grades obtained at A/AS level. The scores for A-level ranged between 2 points (E grade) to 10 points (A grade) and for AS level, the scores ranged between 1 (E grade) to 5 (A grade). The points excluded grades obtained for general studies and for multiple entries in the same subject.

Methodology

Multilevel models; the levels considered were student level; cohort level; establishment level and LA level. Four models were considered.

Model 1 was used the find how the variance in the data is shared between the four levels; this was student 76%; establishment 22%; LA 2%; cohort 0%. Thus the variation in the data is mostly at the student level.

In Model 2 the explanatory variables gender, age, establishment type, region and time were introduced to the model. This model showed females to score about 0.4 of an A level grade higher than males of the same age and establishment type.

Model 3 included GCSE prior performance as well as the explanatory variables of model 2 and showed significantly different results. Female students score about 0.3 of a grade less than males with equivalent GCSE performance. Older students obtain about 0.5 of an A-level grade less than younger students. The effects of most types of establishment are much reduced but selective schools (LA and independent) had positive effects on progress and comprehensive schools had a negative effect.

In Model 4 compositional effects for the establishments were introduced. These were the mean GCSE score and its standard deviation, and the mean number of GCSEs taken. With these variables included independent selective schools and sixth form colleges were seen to have a positive effect.

Results and Conclusions

The authors found that students in maintained comprehensive schools made similar progress to those in state selective schools, when matched on average performance at GCSE of all their candidates. Students in sixth form colleges did better by about 1 A-level grade. Students with lower performance at GCSE made better progress in the independent schools.

When the academic composition of an establishment as measured by average performance at GCSE was left out of the model, then students in maintained selective schools scored on average about 6 score points more than students of the same age and gender in comprehensive schools. Thus selective schools were seen to do better when attainment at A/AS was measured, rather than progress from GCSE.

Males tended to do better then females. Females with grades averaging C or above tended to enter fewer A/AS levels than similarly qualified males, and to obtain lower overall scores. This tendency increased as GCSE score increased.

Older students tended to make less progress from GCSE to A/AS level than younger students. This tendency increased as GCSE score increased.

Students in establishments where the average GCSE attainment was high tended to make better overall progress to A/AS level than those in other establishments of the same type.

Students in the North of England tended to score less highly at A/AS level than students in the rest of England whose performance at GCSE was similar.

Limitations

The authors themselves note several limitations to their study.

No A*=8 in the GCSE scores as this was not introduced until 1994.

Scores do not distinguish between different subjects at GCSE and at A/AS and so does not consider subject difficulty.

The type of establishment where GCSEs were obtained was not considered, nor were vocational qualifications such as GNVQ.

There was no consideration of socio economic background of students, their mobility nor the location and social environment of their A/AS level establishment.

We note that this study used robust statistical techniques and a large data set but the authors have acknowledged there are limitations. However, we agree with the conclusion that average GCSE performance of students in an establishment is a significant predictor of individual progress. It is notable that these authors did not set out to compare selective and non-selective education systems as a whole, but different types of post 16 establishment, and the gender and age of students who attended them. We also note how the choice of combination of explanatory variables in the model affects the outcome, particularly in relation to the academic composition variables and how this reduced the effect of other variables. The role of compositional variables is an area in need of further study and we discuss this further in Chapters 8 and 9.

3.2.7. Schagen and Schagen (2003): Analysis of National Value-added Datasets to Assess the Impact of Selection on Pupil Performance

Schagen and Schagen used nationally available data sets and multilevel modelling techniques to investigate whether in maintained LA schools, comprehensive or selective education systems produce the best overall results. They did this in two stages being limited by the pupil-matched data actually available; from Key Stage 2 (1995) to Key Stage 3 (1998), and then to GCSE results (2000). They used a value added approach in which prior

attainment at Key Stage 2 was related to performance at Key Stage 3, and then Key Stage 3 was taken as the measure of prior attainment and related to performance at GCSE.

The question they were investigating was to what extent is performance influenced by the type of school attended and the selection policy in the LA in which the school is situated. They classified LAs as having high selection (30% or more of pupils attend a grammar school), moderate selection (20% to 29%) and low selection (below 20%). They focussed principally on high and moderate levels of selection and within these LAs termed all non-selective schools to be secondary modern schools.

They addressed the question in two parts:

1. Does comprehensive or selective education produce the best overall results?
2. Are different types of secondary school better for different groups of children?

Data

National data sets obtained from the DfES. This data comprised information on pupils and their performance at key stages, namely Key Stage 2 1997 to Key Stage 3 2000 (387595 pupils, 3034 schools, 149 LAs), and Key Stage 3 1998 to GCSE 2000 (482399 pupils, 3124 schools, 149 LAs). These were the most recently available national data sets at the time.

Explanatory variables

Gender, type of school (grammar or not), size of year group in school, percentage taking free school meals, percentage in LA attending grammar school, prior achievement indicator.

For Key Stage 2 to Key Stage 3, the prior achievement indicator was level achieved at Key Stage 2 in mathematics, English and science. For Key Stage 3 to GCSE the prior achievement indicator was the level achieved in mathematics, English and science.

Outcome variables

At Key Stage 3 the level achieved in mathematics, English and science

grades achieved in each of the subjects at GCSE or GNVQ. This was total GCSE points score (with A*=8, A=7 …G=1) average GCSE score, and score in mathematics, English and double science.

Methodology

Multiple regression; multilevel modelling and logistic regression

Results and Conclusions

Key Stage 3 1998 to GCSE 2000

For pupils with KS3 performance at about level 5 and level 6 or 7, the authors found a small difference in outcome at GCSE in favour of pupils who attended grammar schools. Very high attaining pupils at KS3, said to be about level 7 and above, did not show this difference.

LAs where over 20% of pupils attend grammar schools are associated with slightly less progress in either grammar schools or comprehensive schools, than is the case where the selection level is lower.

In comparing comprehensive and selective systems (selection at 20% and over) they found the differences in progress were not particularly striking, though it was a little higher for comprehensive systems, especially at higher levels of attainment.

Key Stage 2 1995 to Key Stage 3 1998

The authors found a distinct difference in KS3 test levels for pupils on the 'borderline' (their quotes) at KS2 between grammar schools and secondary modern schools.

Taking 'borderline' as an average KS2 level of 4.5, pupils in grammar schools are about half a level higher when they reach KS3, compared to those in comprehensive and secondary modern schools. This can be interpreted as the pupils being about one year ahead in the grammar schools.

For pupils with prior attainment at about level 4 or 5, pupils in grammar schools make more progress than those in comprehensive schools. This difference is not apparent for pupils who have higher attainment at KS2, ie, above about level 5.

The authors investigated whether being at a grammar school enhanced the chance of a pupil with a given ability being entered for a higher tier at Key Stage 3. They found that in mathematics the chance of being entered for a higher tier is 9 to 16 times higher in a grammar school compared to a comprehensive school, and 18 to 20 times higher for science.

The Grammar School Effect in Key Stage 3

Schagen and Schagen discuss what they call the grammar school effect in KS3, that is why pupils of equal ability as measured by KS2 prior attainment make more progress in grammar schools. Possible explanations were put forward.

1. The number of pupils entitled to free school meals is smaller in grammar schools. Percentage of pupils on free school meals is known to affect value added performance.
2. KS2 level is too blunt an instrument. They state that pupils entering grammar school might be of higher aptitude (we note they chose the word aptitude here, not ability) than those who were not selected. It may be that other aspects on which selection decisions are made, such as teacher's recommendation, and entrance tests are more reliable as predictors of later achievement. Pupils who enter grammar school may be more able (we note they now use this term) than others with similar attainment at KS2, although other research by the authors did not confirm this possibility.
3. In selective areas the importance of KS tests might be down played relative to selection tests for the grammar schools. The authors dismiss this possibility on the grounds that a similar progress effect would be expected in non selected pupils at KS3 and this is not apparent, and in grammar schools there is no apparent reason why the effect should be highest for the borderline pupils.

4. Motivated pupils who attend grammar schools, supported by motivated parents might make more progress than suggested by their KS2 prior attainment. The authors noted that percentage free school meals is a blunt indicator of deprivation, and will not differentiate for pupils family backgrounds.
5. The structure and environment provided by a grammar school may enhance pupil achievement. To investigate this, the authors compared the tiers of entry at KS 3 for mathematics and science for grammar school and non-selective pupils. They found that entry to higher tiers was much more likely in the grammar schools.

We agree with Schagen and Schagen in that what they achieved in this research demonstrated the potential of the national data sets and multilevel modelling to investigate questions such as the effect of selection at age 11 in some parts of the country. The authors are tentative in their conclusions (far more so than Jesson) and emphasise the need for further research.

In particular they have highlighted the difficulty in measuring an attribute in a 10 year old child called ability or aptitude, and then using a measure of that as a baseline with which to judge progress made at a later date.

Their work can be criticised for some of the same problems as that of Jesson. Jesson made the assumption that schools in the LAs he deemed selective that were not grammar schools were secondary modern schools. Schagen and Schagen do the same, although as noted previously many of these schools are deemed to be comprehensives by the DfEE. We note that the present authors focussed on a selection level of 20% and above so their results would not necessarily apply to areas with a lower level of selection.

The other major criticism is that of value added between KS3 and GCSE which we previously argued (following Prais 2001) was difficult, if not impossible, for those pupils who had high achievement at KS3 to demonstrate at GCSE.

Methodology: it is difficult to be critical of the methodology in that in setting up a multilevel model decisions have to be made about what outcomes to define as a result of the schooling process, and what the input factors that may affect those outcomes should be (we discuss this further in our own modelling in Chapter 8). The authors' assumptions are justified by the high level of the variation in the outcome measures that they are able to explain.

In the two questions these authors set out to answer, the terminology itself is open to debate. What is meant by best results overall? The authors take it as a compromise between the progress seen in value added between KS2 and KS3 and then again between KS3 and GCSE. The measurement of results is itself limited. As mentioned before Prais noted the limitations of KS3 as based on only mathematics, English and science, (as is KS2) and measures of GCSE attainment in total or average point scores hides the subjects they were taken in, and it is assumed these are all of equal difficulty. Another aspect of looking only at results at GCSE is what happened next for these pupils. What did they go on to achieve in post 16 and higher education and ultimately their careers? This aspect is worthy of further research.

The authors also refer to different kinds of secondary school and different groups of children. Their different kinds of school were essentially grammar schools and non-selective schools, although they gave some discussion in which the latter category was broken down into comprehensive and

secondary modern schools. In 2005 there are many different types of school (Foundation; Faith, Voluntary Aided, Community and so on – the authors pursue this in their 2005 paper) as well as the specialist schools. Similarly, to classify children in different groups by the KS2 levels they attain is crude. However, the issue raised by Schagen and Schagen that their borderline children who were not selected were disadvantaged in that they made less progress than those that were selected, is worthy of further research. Certainly, it would be important to rule out 'regression to the mean' as an explanation for this apparent difference. They also raise the question of whether such pupils underachieve because they view themselves as failures for not being selected, though this is somewhat speculative.

The authors' final conclusion is that they have shown children in a certain ability group can perform above expectations in a particular school context. They say this raises a challenge for all schools, but we interpret this very much as the environment and expectations of a grammar school as raising their pupils' performance.

3.2.8. Schagen and Schagen (2005): Combining multilevel analysis with national value-added data sets – a case study to explore the effects of school diversity.

This is a follow up study to the authors' 2003 paper, using a more recent data set. They were able to match 380,000 pupils from their Key Stage 2 levels in 1996 to their GCSE results in 2001 and so consider progress as measured by value added over 5 years of secondary schooling. Rather than comparing selective and non-selective LAs they compared grammar schools to non selective schools without reference to actual LAs. They also extended their research to other types of school, particularly faith schools and specialist schools. The authors again emphasised that the availability of national data sets together with developing techniques in multilevel modelling would give educational researchers new opportunities to carry out objective research into issues of real importance, which have hitherto been prone to speculation and inconclusive argument from entrenched positions.

Data

National value added data set from Key Stage 2 in 1996 to GCSE in 2001, with matched data for about 380,000 pupils in about 3000 schools, including 156 grammar schools.

Explanatory variables

Pupil level: gender, level achieved at KS2 in mathematics, English and science, an indicator to show if average KS 2 level was below 3.

School level: percentage of pupils eligible for free school meals, and the square of this (research has shown this is a better indicator); grammar school indicator; indicators for types of specialist school and faith school.

LA level: percentage of pupils in LA attending grammar schools and attending specialist and faith schools.

Outcome variable

Points score at GCSE based on A*=8, A=7 down to G=1. This was detailed as total points score, average points score, score in mathematics and English,

total science score and average science score, and the number of GCSE entries.

Methodology

Multilevel modelling with levels at the pupil level, school level and LA level.

Results and Conclusions

For the comparison between grammar schools and comprehensive schools the authors largely confirmed their previous results. They noted a large grammar school effect, in which borderline pupils – those who narrowly obtain a grammar school place – obtain much better GCSE results five years later than pupils of equal prior attainment in comprehensive schools. However pupils with KS2 average level higher than 5 may fare better in comprehensive schools.

Boys outperformed girls in GCSE mathematics and science, but girls outperformed boys in all other outcomes considered (total point and average point score, number of GCSE entries, GCSE English)

GCSE performance relative to KS2 attainment tends to be reduced in schools with high numbers of pupils on free school meals. The impact of this though falls off after about 50 to 60% FSM.

The authors noted they were considering value added and not raw scores, but still the grammar schools had far the biggest impact by school type. They noted it would be expected that grammar schools would have the best raw score results.

In grammar schools the difference in range of pupil's performance at KS2 is significantly reduced by the time they take GCSE. Borderline pupils (as defined in their 2003 paper) with KS2 average level at about 4 to 5 perform better in grammar schools as shown by total point score than in comprehensive schools, whereas the reverse is true for pupils with higher KS2 average levels. This was also seen in the other outcome measures except for science, where the grammar school performance was best. The authors attributed this to pupils likely being able to take three separate science subjects in a grammar school as opposed to double science.

Grammar school pupils tend to enter their pupils for the same number of GCSEs, about 9 or 10, whereas comprehensive pupils whose average level at KS2 was about 6, are entered for about 11 or 12 GCSEs.

At LA level, they again classified LAs as low selection and high selection (but did not define these selection levels) and compared them with fully comprehensive LAs. They found in all but one of their seven outcome measures that schools in selective LAs obtained better value added results than in those in comprehensive LAs, but none of the results were statistically significant (i.e. they could well have happened by chance)

They state: selection has no significant impact on the overall performance of LAs and these is little to choose between LAs in terms of overall attainment.

The grammar school effect identified in the 2003 paper was confirmed. They stated their new analysis demonstrates clearly that grammar schools enhance the performance of their least able pupils, raising it to a level well above that which pupils of the same ability in comprehensive schools would achieve. They raise the issue of pupils of essentially the same ability at KS2 of having

rather different outcomes at GCSE depending on what type of school they attend.

They also state this analysis indicates that selection has no significant impact on the overall performance of LAs.

Limitations

Schagen and Schagen point out the assumption of linearity in value added. That is for instance that improving from grade G to F (or 1 point to 2 points) is equivalent to improving from A to A* (or 7 points to 8 points) and this assumption is inherent in their work, and all work, on value added. This assumption is present in using GCSE total or average point score as an outcome measure and they acknowledge that no measure, or indicator, can be perfect. We also note that these authors describe 5 A* to C grades at GCSE as a poor indicator to use as it can be seriously influenced by the performance of a few borderline pupils, and does not reflect any changes in those at the top and bottom of the attainment spectrum. They therefore rejected this measure and used measures which are averages over all pupils.

We question the authors' suggestion that high attaining pupils at KS2 do better in comprehensive schools in terms of total point score, as these pupils tend to take more GCSEs. It seems likely that if grammar schools were to enter such pupils for a comparable number of subjects their total point scores would also be higher.

We also note that their result for grammar school 'borderline' pupils represents a somewhat uncharacteristic comparison, since a typical pupil with a mixture of level 4s and 5s at KS2 would be unlikely to gain entry into a grammar school; those who do could well be significantly different in other ways. Also the number of pupils who are even entered for level 6 papers at KS2 is very small so any conclusions based on this group (i.e. those with average KS2 level above 5) must be treated with some caution. In our own analysis of over half a million pupils in a year group, only about 1200 had an average KS2 score above 5.

All the criticism about the difficulty of defining selective LAs and distinguishing comprehensive schools from secondary modern schools in such LAs (as discussed for Jesson and Schagen and Schagen 2003) can be reiterated here.

It is also interesting that the issue of the grammar school effect and borderline pupils that the authors raise in their conclusions is not discussed further as in the 2003 paper. This essentially is that pupils of apparently the same ability as measured by KS2 average level perform differently in selective and comprehensive schools. Again, we must stress that KS2 levels seem to be too broad of measure to make a clear distinction here.

3.2.9. Atkinson, Gregg and McConnell (2004): The result of 11 plus selection; An Investigation into Equity and Efficiency of Outcomes for Pupils in Selective LEAs.

The research reported in this paper sets out to inform the debate about grammar schools through assessing the impact of selection at age 11 in those LEAs which still have a significant number of grammar schools. This was deemed to be the 19 LEAs where 10 per cent, or more, of the pupils attend a grammar school for their secondary education. The authors investigated

whether these LEAs achieve better GCSE results for their pupils compared to non-selective LEAs. To do this they matched these LEAs to similar parts of the country where there is no selection at age 11. They also investigated which pupils benefit from academic selection. Although they found that selective LEAs do not overall achieve substantially better performance, the authors found that those pupils who did attend a grammar school achieved about four GCSE grade points higher totalled across all subjects than equivalent pupils in the non-selective areas. Pupils who do not attend grammar schools in selective LEAs do slightly less well than comparable children in non-selective LEAs by about one GCSE grade point.

The authors also considered the impact of single sex and religious schools.

Data

The authors used the national dataset, the PLASC data (pupil level annual school census) for 1997 to 2002, and thus tracked the pupils from entry at age 11 to GCSE five years later.

Explanatory variables

Selective LEA, gender, age, school size, free school meals, special needs, English as a second language, ethnicity, single sex school; FSM (free school meals). KS2 score (given numerically as levels 1 to11 without further detail)

Outcome variables

Total GCSE point scores; best 8 GCSE point scores where grade A*=8, A=7 and G=1. Value added between Key Stage 2 and GCSE.

Methodology

OLS Multiple Regression

Results and conclusions

Pupils who attended a grammar school achieved about four GCSE grade points higher than equivalent pupils in the non-selective areas. Pupils who did not attend grammar schools in selective LEAs did slightly less well than comparable children in non-selective LEAs by about one GCSE grade point.

Selective LEAs did not overall achieve substantially better performance than non-selective LEAs in similar parts of the country.

Using free school meals as a measure of poverty, the authors found very few poor children in grammar schools. The small minority of children from poor backgrounds who did attend a grammar school achieved highly, but the authors state that there were other equally able children from such backgrounds in selective LEAs who do not attend a grammar school. The measure of ability was Key Stage 2 results at age 11. Poorer pupils in selective LEAs are only half as likely to attend a grammar school as those with similar Key Stage 2 scores. The authors concluded that if access could be widened then the case for keeping selective education would be enhanced.

More girls than boys attended grammar schools although the split was close to 50%.

Pupils who attend the grammar schools were on average slightly older, but by only about 10 days.

The authors also stated that pupils with special needs and English as a second language are under-represented in grammar schools.

Further detail

FSM (free school meals) and grammar school places

The authors quote figures for the 19 LEAs[2] they included in their study, as follows: 12% of pupils in non grammar schools are eligible for FSM, but in the grammar schools this is 2%. It was noted that non-grammar schools included both secondary modern schools and comprehensive schools. Put another way, of the children who are eligible for FSM in these selective LEAs, 5.8% attend a grammar school. For the children not eligible for FSM, 26.4% attend a grammar school. For top ability children at age 11 as defined by Key Stage 2 score, 32% of those eligible for FSM attend a grammar school, compared to 60% for those not eligible for FSM.

Attainment and Selection

The authors take it as given that the selective LEAs in their study have characteristics that set them apart from the rest of England. These characteristics included affluence as measured by lack of Free School Meals, a high population density and not being a county authority. They associated this with the number of Conservative Councillors in the Authority on the argument that Conservative areas are more likely to be affluent and to have retained selective education. The authors claim their matching appears to be successful based on progress in primary schools between Key Stages 1 and 2 where the slight gap in favour of selective LEAs when compared to all others is substantially reduced when compared to their matched LEAs.

The authors consider various multiple regression models, introducing the explanatory variables in batches. When controlling only for gender, age and school size the Selective LEA effect is 3.6 GCSE grades, but this reduces to 0.7 when all variables are introduced. However, single sex schools are seen as raising the GCSE points by about 7 points. The results for capped best 8 GCSEs are similar, but a little smaller at 2.8 points and 0.6 points respectively. The details of how value added between Key Stage 2 and GCSE have been calculated are not given, but the authors give results showing the Selective LEA effect now down to 1.5 and 0.7 GCSE grade points respectively and the effect of single sex schools is much reduced.

The authors repeat the above regressions using only their matched LEAs and the effects become smaller and even become negative in one model.

The results of theses regression models are summarised in Table 20.

Table 22 shows the regression coefficient, measured in GCSE grade points for the effect of the Selective Area for the authors' four models.

Models and explanatory variables:

> Model 1: selective LEA, gender, age, school size
>
> Model 2: as Model 1 plus FSM, special needs, mother tongue not English, ethnicity

[2] Barnet, Bexley, Bournemouth, Buckinghamshire, Calderdale, Gloucestershire, Kent, Kingston, Lincolnshire, Medway, Plymouth, Poole, Reading, Slough, Southend-on-Sea, Sutton, Trafford, Torbay, Wirral

Model 3: as Model 2 plus single sex school

Model 4: as Model 3 plus school percentages for FSM, special needs, mother tongue not English, ethnicity

Table 22: The regression coefficient, measured in GCSE grade points, for the effect of the Selective Area for each of four models.

GCSE outcomes	Model 1		Model 2		Model 3		Model 4	
	All	Matched	All	Matched	All	Matched	All	Matched
Total GCSE score	3.6	1.8	3.1	1.6	1.6	-0.8	0.7	0.3
Capped GCSE score	2.8	1.3	2.4	1.2*	1.3	-0.6	0.6	0.3
Value added (KS2 to GCSE)	1.5	0.8	1.5	0.9	1.0	0.1	0.7	0.5

Notably in the authors' full results for value added between Key Stage 2 and GCSE, the effects of school size, single sex and religious school have more effect than selective LEA. In Model 4 these were approximately 2 grade points for school size, 5 each for boys and girls schools, and 1 point for a religious school compared to 0.3 for selective school.

The authors carried out a further analysis comparing the value added for grammar and non grammar schools in the selective and matched LEAs. These results show the pupils in grammar schools to be achieving substantially better results than pupils of similar ability in non-selective areas with similar populations. The pupils in the non-selective schools in the selective areas do slightly worse than the equivalent pupils in the non-selective areas but these latter results are small and statistically are borderline. Controlling for percent of pupils eligible for FSM reduced these effects from –0.7 to –0.2 whilst reducing the grammar school effect from 4.5 to 3.6.

The authors also considered borderline pupils, those with similar Key Stage 2 scores who attended or did not attend grammar school. They implied they had sufficient detail in the Key Stage 2 results to be able to describe some children at age 11 as fractionally more able than others. The authors' results show that attending a grammar school is even more important for these borderline pupils where the grammar school effect is substantial (7.1 points in model 1 and 4.8 points in Model 2) with little difference in the effect of those who attend a non-grammar school (-3.2 points and –0.2 points respectively). The authors conclude:

> Within selective LEAs making it to a grammar school is extremely important but those failing to make it are only marginally underachieving those with similar ability in the matched non-selective areas. This high achievement is focussed especially on borderline students who make it into grammar schools.

The authors carried out a further investigation into the effects of grammar schools and non grammar schools on children from poor backgrounds, by comparing the value added for those eligible for Free School Meals and those not. They again found the grammar school effect to be substantial at about 7

to 8 grade points. The effect on those not eligible for FSM was similar to grammar schools as a whole, at around 3 to 5 grade points. Thus they conclude the small number of FSM pupils who attend a grammar school achieve highly.

We note this research also highlights the issue of borderline pupils. The measure of ability at age 11 is Key Stage 2 results yet the discussion of borderline pupils suggests that this is too broad a measure to distinguish the academic potential of such children. It seems likely that an entrance test that gives a measure of cognitive ability does differentiate between these borderline pupils as the evidence is that borderline pupils, defined by Key Stage 2 score, do not do as well in non-grammar schools as others do in grammar schools. Selection tests identify potential whereas the Key Stage 2 results are a measure of what has been learnt up to age 11 in mathematics, English and science. However, most selection tests and their results are confidential to the school concerned, and without access to them, it is difficult to pursue this further. It is also not clear in this paper just what Key Stage 2 results are being used; they are clearly not levels as defined by QCA, but what they are is not discussed. Once again, it is really impossible to rule out a regression-to-the-mean effect as the true explanation of this apparent phenomenon.

The authors state the LEAs considered, i.e. those with over 10% selection, are more politically Conservative and affluent than other parts of the country. Although we agree with the LEAs concerned (Table 1, p12) and note these do not include Birmingham, we also note the authors have offered no evidence to back this statement, yet they use the number of Conservative Councillors in the Local Authority as the match for *similar* parts of the country. Although we would say this is not an unreasonable way to achieve a matching, we have noted the variation in percentage of selection across the country, and have put forward the view that comparison of *similar* areas is fraught with the difficulty of identifying such areas. We note the authors state that similar areas would not be county LEAs but have Lincolnshire, Gloucestershire, Buckinghamshire and Kent in their 19 LEAs. Their claim that the match works are based on very small differences in value added between Key Stages 1 and 2.

The authors accept that there are important explanatory variables, like 11+ score, and social characteristics and family circumstances, which could not be included in their models as measures were not available.

The authors also note that the extent to which pupils cross LEA boundaries to go to school has not been taken into consideration. This figure would appear to be vary considerably across the country. These authors quote Schagen and Schagen in a study of Slough, where they found 40% of the grammar school pupils in Slough come from outside the Authority. In 2002 the DfES quoted a national figure of 94.7% as going to school in the LEA in which they live. In our own analysis we found this to be 92%, and in the grammar schools in Birmingham to be 75%. It is also notable that having decided to use data from the 19 LEAs they deem to be selective, they then aggregate these as if they were homogenous averaging about 25% selection. However, as we have already noted, the level of selection varies considerably across these LEAs so the results from this paper can only be indicative.

It is also notable that the authors make an issue of under-representation of Free School Meal children. FSM is a crude indicator of poverty and it would

be a suitable area of further research to investigate the relationship of achievement at school to more detailed deprivation indices. We must also remember that there is a general relationship between socio-economic status and measured ability, so if grammar schools select the most able pupils at 11+, it would be unreasonable to expect them to have a representative proportion of those eligible for Free School Meals. The fact that they do not is not necessarily evidence of bias in the selection process. We return to this question in Section 6.1.2.

Atkinson, Gregg and McConnell later (in 2006) reviewed their 2004 study and reached the same conclusions. The only addition to their initial study was an examination of any potential biases in their estimates of the grammar school effect using an instrumental variables approach. This specifically explored the variation in pupil attainment from attending a grammar school that is derived from age within the school year rather than ability.

Studies by Burgess *et al* (2004, 2005) and West & Hind (2006) have looked at the impact of choice, sorting and selection of students in selective and non-selective schools. These are reviewed in the following sections.

3.2.10. Burgess, McConnell, Propper, Wilson (2004). Sorting and Choice in English Secondary Schools

Burgess *et al* analyse student-level data from England to investigate the impact of choice on pupils' sorting or segregation across schools. They focus on two different routes through which pupils are assigned into schools: the 'elite schooling assignment' mechanism through which pupils are allocated into grammar schools on the basis of their test scores, and the 'neighbourhood schooling' which allows pupils to be allocated into schools according to home-school distance.

The authors used the PLASC dataset which, they state, covers approximately half a million primary and secondary pupils in England. At pupil level, it provides test score histories in addition to some individual characteristics, such as gender and within-year age, ethnicity, eligibility for free school meals and special educational needs. The characteristics at school level include performance measures, geographical co-ordinates, school size, age range, religious denomination, funding status, gender mix and admissions policy. It is not clear, however, whether all of these pupil and school level characteristics have been used in this study apart from the home postcode of every pupil being explicitly mentioned as it allows each pupil's location in relation to school. Pupils' KS2 scores are used to analyse their sorting into secondary schools. Mostly state secondary schools are used in the analysis with other types of schools (such as private or special schools) having been omitted as KS2 results are not provided for all private schools. However, it is argued that, as these schools are of interest, some relevant evidence will be presented. They use both selective (defined as one in which more than 10% of pupils attend a grammar school) and non-selective LEAs in the analysis. The authors do not specify though which LEAs and how many of them were included in their analysis nor the process through which these LEAs were selected.

Using the co-ordinates for each school, the authors construct drive-time zones (DTZs) around schools. They use a 10-minute DTZ to count the number of 'nearby' schools as their measure of the extent of school choice. They then aggregate this to LEA level to draw comparisons with their measures of

sorting and segregation. Student sorting is measured across ability, ethnicity and disadvantage. Segregation is measured across schools and electoral wards taking the LEA as the aggregate. The authors use as their measure of segregation the 'dissimilarity index' which 'is based on the idea of segregation as unevenness of the distribution of different types of students across units within the aggregate area; the more uneven the spread, the higher the degree of segregation'.

Results

The authors present their results under the following four sub-sections:

a. Feasibility and Exercise of Choice

They first claim to have established that most students in England are able to exercise choice between different schools. Using the ten minute drive-time zone, they find that secondary schools in England have, on average, more than 6 schools within ten minutes drive of themselves. They find that in the three area types considered, the mean is 17 in the dense London area, 7 in non-London urban and just over 1 in non-London rural areas. They then explore whether this 'feasibility of choice' is actually taken up. They find that about 45% of students attend their nearest school (a striking finding according to the authors suggesting the importance of post-residential school choice) and thus conclude that 'just over half are "exercising choice" in the sense of not going to the closest school, given their place of residence'. However, the authors claim that this may be a forced choice if the local school is filled up.

b. Characterisation of Ability &

c. Characterisation of Multi-dimensional Sorting

The authors claim that ability is one of the main issues in the analysis of choice. They, therefore, attempt to characterise ability sorting and show a correlation between ability sorting at school level and sorting at school level by income and ethnicity. In all three dimensions of ability, socio-economic status and ethnicity, there is variation in the degree of student sorting across the country, but particularly among high ability students. Those areas that have retained a selective system show the highest degree of sorting. According to Burgess *et al* (2005), this finding demonstrates that 'selection through the housing market has not simply replicated grammar school patterns'.

d. Post-residential Choice and Sorting

The authors then measure the degree of choice as the average number of schools that can be reached within a particular drive time and find that school sorting is higher compared to neighbourhood sorting in areas where there is more choice in the number of schools. In other words, 'markets in which there are more schools are markets in which there is more ability sorting'. They also show that there is a difference in the relationship between neighbourhood sorting and school sorting in selective and non-selective LEAs. The selective LEAs show high levels of high ability sorting whereas they do not have high levels of neighbourhood sorting of ability. This demonstrates, according to the authors, 'the divorce of residence choice from school choice that arises in elite schooling areas'.

3.2.11. *Burgess, S., Propper, C. & Wilson, D. (2005): Will More Choice Improve Outcomes in Education and Health Care? The Evidence from Economic Research. The Centre for Market and Public Organisation*

Burgess, Propper and Wilson summarise the evidence from economic research to shed light on whether more choice can improve outcomes in education and health care. They argue that 'consumer choice acts as a major driver for efficiency' and that more choice given to service users, in this case as far as schools and hospitals are concerned, will result in a continual drive for improvement in an attempt to attract consumers.

From the evidence on economic research, the authors conclude the following:

1. A choice policy implemented successfully would improve standards for most school students as the competitive pressure introduced by choice will push up standards. In addition, a move to a choice system would result in a fall in house prices indicating gains for the working class. This outcome would allow some groups, in particular, to do well, for instance poor children or those from ethnic minority backgrounds who would not be able live near a good school to secure a place there. The authors specifically point out that it may well be the pressures of competition rather than the actual exercise of choice that determine the main impact of choice. Most of the evidence comes from the US which suggests that 'schools facing a competitive threat respond by increasing productivity'. However, even though empirical evidence shows that there are test-score gains for some of the students who exercise choice, this result cannot be generalised when different types of student or different types of choice programme are considered.
2. There will be a need for flexibility in the supply of school places. Policy needs to enable existing schools to expand in order to meet demand, new schools to start and poor schools to close. This flexibility in supply will tackle 'sorting' and segregation of students on the basis of neighbourhood or ability. The authors draw on evidence which suggests that parental choice in addition to flexibility in the supply of school places reduce 'sorting' of students by income and ability. However, parental choice with poor flexibility in the supply of school places leads to increased sorting.
3. The role of peer groups partly determines the effects of school choice. It is argued that the outcome of a choice policy can be more problematic if the consideration of peer groups affects parents' school choice. The effect of choice on improving school quality will be limited if high scores are achieved primarily because of a good student intake. Conversely, 'if peer effects are important, and if sorting increases with choice because of insufficient flexibility on the supply side, then students who find themselves in schools with less able peers will perform less well. This process is likely to cumulate in that poor achievement in one year will attract a less able group of students in the following year, thus compounding the problem'.

3.2.12. *West, A. & Hind, A. (2006): Selectivity, admissions and intakes to 'comprehensive' schools in London*

The focus of this study is on admissions criteria that 'comprehensive' secondary schools in London use. The authors state that London has been chosen as an area of investigation as it presents an interesting study case in terms of its size and particularly its diversity due to the high levels of disadvantage within the maintained school sector. The paper aims to extend previous research by looking at one of the reasons why huge variation has been found to exist between the prior attainment of pupils entering London secondary schools. They specifically look into admissions criteria and practices used in the case of there being more applicants to a school than there were places available.

The study aims to answer the following research questions:

- What admissions criteria and practices are used for London secondary schools, and how do these vary between schools that are responsible for their own admissions and those whose admissions are the responsibility of the local authority?
- Are there differences between the admissions criteria and practices used in London and in the rest of England?
- To the extent that there are differences between school types in terms of their admissions criteria, are these associated with differences in selected school intake characteristics and outcomes?

Admissions criteria were examined by focusing on published criteria/practices used by state-maintained secondary schools in London for entry into Year 7 in the 2001/02 academic year. The authors made comparisons with the rest of England but they state that, as relevant data were not provided by some schools not included in LA prospectuses, the results are not fully representative of all schools in England. A total of 2862 secondary schools in England were included in the final analysis (2023 community schools, 81 voluntary-controlled, 401 voluntary-aided and 357 foundation schools) with 382 of these schools being in London (208 community schools, 6 voluntary-controlled, 107 voluntary-aided and 61 foundation schools).

Results

Regarding the admissions criteria used by secondary schools in London (the authors' first research question), it was found that over nine out of ten secondary schools gave priority to siblings and almost nine out of ten gave priority to distance. The authors mention a number of other criteria that were used by schools, for instance, medical/social needs, special educational needs, religious factors, banding, school being 'first preference', child attending feeder school, pupil/parent interview used, child of employee among others. In terms of differences in admissions between different types of schools, the authors reported the following:

- More community/voluntary-controlled (these schools are not responsible for their admissions) than voluntary-aided or foundation schools (responsible for their own admissions) indicated that priority was given to pupils with medical/social needs and special educational needs.

- One fifth of schools used 'banding' to ensure a 'balanced' ability intake. The inner London area was more likely to use some form of banding; six out of ten secondary schools compared with 1% in outer London.
- Only voluntary-aided schools reported interviewing pupils or parents in order, for instance, to establish parents' religion following guidance on school admissions.
- More voluntary-aided and foundation schools than community/voluntary-controlled schools reported giving priority to a proportion of children with ability or aptitude in a subject area and to children having a strong family connection to the school.
- Finally, more voluntary-aided and foundation schools were found to use at least one criterion/practice which allowed particular types of pupils to be selected, for instance, selecting on the basis of aptitude or ability, or interviewing pupils or parents).

Regarding the authors' second research question, namely, whether admissions criteria varied between London and the rest of England, there were a number of similarities and differences. Criteria that were used in both London and the rest of England were siblings, distance and medical/social need. However, London was more likely to indicate special educational needs, religion, banding, interviews and giving priority to the children of employees and of former pupils as admissions criteria. Further analyses were conducted to find out whether the differences found between admissions criteria in the two areas were associated with whether or not schools were responsible for their own admissions. These revealed that schools in London and more community/voluntary-controlled ones were more likely to report using certain admissions criteria/practices, namely, distance from home to school, medical/social need and special educational needs. More London schools (voluntary-aided) indicated interviewing the child and/or parent. In addition, more schools in London used 'banding'.

After controlling for whether or not schools were responsible for their own admissions, the study found that more schools in London (voluntary-aided and foundation) would use one criterion/practice that was selective or that would be potentially 'creaming' certain pupils. However, schools outside London were more likely to use other admissions criteria, such as catchment areas, attendance at a feeder primary school and the secondary school having being identified as the parents' first choice.

A final analysis was carried out to identify whether any systematic relationships existed between London schools that were or were not in control of their admissions as far as their intakes and examination outcomes were concerned. Drawing on a sub-sample of schools in this analysis, it was found that a higher proportion of pupils with special educational needs attend community/voluntary-controlled schools than voluntary-aided and foundation schools (the differences between the two types of schools were statistically significant). In addition, there was a statistically significant higher percentage of pupils achieving five or more A*-C grade GCSEs in voluntary-aided and foundation schools.

The authors concluded that those secondary schools that are responsible for their own admissions were found to have fewer pupils with special educational needs and performed better achieving 'higher positions in the

published examination league tables'. Their conclusions are in line with the recent debate about some state schools that control their own admissions being highly unrepresentative socially and often academically of their surrounding area and about the need for school admissions to offer parents fairer choices (Education Guardian, 2005; IPPR, 2007).

A limitation of their study though, which they acknowledge, is that, as data regarding pupils' attainment when entering the secondary schools were not available, the eventual differences in attainment may be due to voluntary-aided or foundation schools being more 'effective' than community/voluntary-controlled schools. However, they maintain that, according to their findings, 'there may well be creaming or selectivity by schools that are responsible for their own admissions… children with special educational needs appear to be disadvantaged in London in relation to gaining access to voluntary-aided and foundation secondary schools'. They argue though that the answer may be that parents are reluctant to apply to these schools as they may not expect to be offered a place or may feel that the schools will not cater effectively for their child's needs. The authors support the view that school autonomy in admissions needs to be limited in order to restrain the 'current hierarchy of schools in London'.

3.2.13. *Levačić and Marsh (2007): Secondary modern schools: are their pupils disadvantaged?*

In this piece of research, Levačić and Marsh aim to investigate the performance and resourcing of secondary modern schools in England focusing on the ways in which pupils attending these schools may be disadvantaged. The analysis is carried out using data from English national data sets in addition to a survey of school costs undertaken on behalf of the Buckinghamshire Upper Schools Forum which is working towards increasing local authority funding for upper schools. The authors attempt to address the gap in the literature on the funding of secondary modern schools in relation to other types of schools. They use information from the DfES Annual Schools Census dataset (2001) and Ofsted (2002) to argue that secondary modern schools have larger numbers of socially and educationally disadvantaged pupils (a higher proportion of students in these schools have special educational needs and are more likely to be permanently excluded).

The analysis consists of two main parts. In the first part, a national data set is used to match pupils' performance at Key Stage 2 and GCSE. The second part involves an examination of the financing of secondary schools in England.

Data

The authors used the QCA national data set at pupil level to conduct a value-added analysis of GCSE results in 2001 matched with prior attainment data for Key Stage 2 in 1996. The data set consists of the National Curriculum test and examination results in addition to the age, gender and school of over 330,000 state school pupils (independent, middle and special schools are excluded).

Explanatory variables

Pupil-level data: Key Stage 2 score (average marks for English, mathematics and science), age and gender.

School-level data: Selective LEA, type of selective school, average percentage of pupils eligible for Free School Meals, those with SEN statements, white pupils and pupil-teacher ratio (average for 1997-2001).

Outcome variables

Total GCSE/GNVQ points score where grade A*=8, A=7 and G=1 and the probability of obtaining 5 or more A*-C GCSE/GNVQ passes. Value added between Key Stage 2 and GCSE.

Methodology

Regression analysis; Multi-level modelling

Results and conclusions

They found that GCSE grades were slightly lower for pupils at secondary modern compared to comprehensive schools whereas grammar school pupils' attainment was considerably higher. The effect of attending a grammar school was 5.5 additional GCSE/GNVQ points compared to a comprehensive school whereas the average estimated effect of attending a secondary modern was 1 grade less at GCSE in comparison to being at a comprehensive school. The analysis also indicated that, as the ability of the pupil increased, the advantage of attending a grammar school compared to a comprehensive decreased. The results also showed that for more able pupils as shown by their attainment at KS2, the loss in GCSE points is greater from attending a secondary modern school.

Additionally, an average ability pupil (boy or girl) had a 30% higher probability of gaining 5 or more grades at A*-C at GCSE (considered here as another measure of GCSE attainment) when attending a grammar school in comparison to a comprehensive and between 3 and 4% less chance in comparison to a secondary modern. Regarding pupils of higher ability (1 standard deviation above average), however, boys attending grammar school had a 15% better chance of achieving 5 good GCSEs and girls a 10% chance. Both girls and boys were 2-3% less likely to achieve 5 good GCSEs if at a secondary modern. The results indicate, therefore, that pupils of average ability are the ones that gain most by attending grammar school. This is consistent with Schagen and Schagen's (2003) findings.

The authors agree with previous research that it is difficult to determine whether the effect of grammar schools on attainment is the result of the positive influence of the peer group or whether it is the schools' focused academic mission that leads to these attainment gains. They conducted, therefore, a further analysis to shed light on the attainment disadvantage of secondary modern schools by showing that these schools have higher numbers of socially disadvantaged students – 4% more pupils eligible for Free School Meals which results in an average reduction in attainment for pupils attending secondary moderns by 0.6 GCSE grades. However, the gains of a better peer group for those pupils attending grammar school were estimated to be 1.8 additional grades.

A further analysis was conducted in relation to how secondary modern schools are funded. Data used for this analysis included the school-level variables in addition to two measures of revenue per pupil per annum: budget share and current revenue per pupil (obtained by the DfES). The authors present some descriptive data to show that there was little difference

in budget share per pupil between secondary modern and grammar schools although there was a slight increase in favour of secondary moderns over the three years considered (2000/01-2002/03). They also indicated that for a number of LEAs examined in 2000/01, the budget share per pupil in a secondary school was less than the budget share per grammar school pupil although variability was found among the 10 selective LEAs considered.

However, as this descriptive analysis does not take into account differences in the factors that determine school revenues, in other words, the 'cost drivers' (e.g. the size of school, the age range of pupils, differences in regional costs, additional learning needs, Standards Fund eligibility, differences in LEAs' estimated need to spend on secondary education, whether the school has got sixth form and the fact that wholly selective LEAs fund less per pupil), the authors subsequently carried out a regression analysis aiming to control for these factors. The results indicated that secondary moderns were funded less per pupil (around £80 less per pupil) compared to comprehensive schools whereas grammar schools were funded more (over £100 more per pupil) - the coefficients providing an estimate of how much extra or less a comprehensive or secondary modern school received per pupil compared to a comprehensive school were statistically significant for all years apart from the year 2002/03 for secondary modern schools.

A further analysis indicated that secondary modern schools are more likely to run budget deficits than comprehensive schools – fifty-four per cent of secondary moderns were found to be in surplus for all three years compared to 64% of comprehensives and 88% of grammar schools, and fifteen per cent were in deficit compared to 94% of comprehensives. The authors maintain that budget deficits have strong negative implications for schools and their students including the fact that a school with a budget deficit is not in a position to apply for specialist school status or it has no fund to use in case of an expenditure emergency. They, therefore, conclude that 'the foregoing evidence on funding and the financial health of schools point quite clearly to secondary modern pupils being on average disadvantaged by lower funding per pupil than pupils in comprehensive schools, and particularly relative to grammar schools'.

An additional survey that was carried out in Buckinghamshire revealed a number of specific need-related activities which would incur additional cost per pupil regarding secondary modern school educational provision. These included vocational courses, special educational needs, pastoral care and recruitment and retention issues as well as additional contributions from the school budget for implementing the Key Stage 3 strategy. The evidence from this questionnaire on costs in addition to the finding that secondary modern schools run greater budget deficits, with a particularly high incidence in Buckinghamshire, indicate, according to the authors, that 'these schools do face additional costs for which they were not adequately funded'.

A final analysis was carried out where home-to-school transport expenditure per pupil at LEA level was associated with a number of factors, i.e. low population density, the LEA's additional educational needs indicator and average school size. The regression analysis showed that low population density had the largest effect on expenditure followed by the proportion of pupils attending grammar schools. It was found that every additional 10% of LEA pupils attending grammar schools roughly corresponds to a £10 increase per pupil in transport costs. The authors claim that these resources that are

spent on transporting pupils to more distant schools should be spent on teaching and learning, and that parents of grammar school pupils need to pay for transport to school in order to reduce the home-to-school transport budget, as parents of pupils attending secondary modern schools do.

In conclusion, the authors state: 'it is clear from the evidence presented that secondary modern schools are under-funded in relation to the needs of their pupils… the selective system in 10-20 local authorities works to disadvantage pupils who end up in secondary modern schools. In a few authorities the resourcing policies actually discriminate against secondary modern pupils compared to grammar school pupils. This is the downside of local political discretion to choose selective education'.

The authors claim that pupils' KS2 total score in the core subjects of English, mathematics and science allows valid comparisons to be made regarding pupil progress in comprehensive, grammar and secondary modern schools. However, we have already mentioned the limitations of KS2 tests in measuring achievement in only those three subjects. We also note that the survey data on costs come from only one education authority which limits any generalisation of the findings. The authors conclude that 'the selective system in 10-20 local authorities works to disadvantage pupils…'. Apart from the 10 authorities specified as being wholly selective, it is not clear which ones are the remaining authorities. The LEAs involved in the analysis of home-to-school transport expenditure are not explicitly mentioned. In addition, the authors accept that there is a possibility of 'an upward bias in the estimated costs' as the purpose of the questionnaire was part of a campaign for additional resources and it could be expected that head teachers would have an incentive to bias upwards the reported costs.

3.2.14. *Maurin, E. & McNally, S. (2007): Educational Effects of Widening Access to the Academic Track: A Natural Experiment.*

In this study, Maurin and McNally make use of a 'natural experiment' to investigate the effects on educational attainment of a reform which widened access to schools providing a more academically orientated education in a particular region (Northern Ireland) at a particular point in time (the size of the cohort entering grammar schools increased from 31% to 35% following the reform). The educational outcomes of the reform are considered using England as the comparison group which acquired a mostly comprehensive status in the 1960s and 1970s whereas Northern Ireland retained its grammar school system. As the authors state, both regions have a similar curriculum and students take the same national examination at the ages of 16 and 18. However, a reform taking place in Northern Ireland resulted in a significant increase in the number of pupils attending the more academic track (grammar schools) at the end of primary school, between the pre-reform birth cohort (children born in 1978) and the post-reform birth cohort (those born in 1979). This change allowed the authors to compare educational outcomes in Northern Ireland and England, before and after the reform looking specifically at the overall effect of allowing entry to the 'academic track' for a number of borderline pupils who would previously not be admitted. Using a 'difference-in-differences' analysis, they found that the increase in the number of pupils entering the more academic track was followed by an increase in GCSEs and A-levels examination results. They specifically indicate that their approach allowed them to identify the

consequences of the reform in the context of a selective educational system showing the positive effects generated by widening access to the more academic track. However, it did not allow any judgement to be made about whether Northern Ireland's 'fully selective system' is better or worse than England's 'fully comprehensive system'.

However, the attractiveness of their experiment, the authors state, is that the 'de-tracking' reform of increasing access to more academically orientated education was the only differential change that took place between the two regions at that specific time and that the reform only modified the intensity of selection rather than modifying the nature of the school system.

The authors made use of administrative data obtained from the Department of Education in both countries to investigate the impact of the reform on numbers of pupils entering grammar schools and the outcomes on the pre-reform and post-reform cohorts. They state that the strength of their data lies on allowing observations to be made on examination outcomes that are free of sampling error by gender, school type and cohort. They developed a conceptual framework within which three different groups of pupils and the combination of three effects are defined allowing the interpretation of the increase in exam performance in Northern Ireland when compared to England.

Firstly, there is the effect on the group of pupils who attend a grammar school after the reform but who, in the absence of the reform, would have attended a non-elite school. Secondly, there is the effect on those pupils who enter a non-elite school of losing more able pupils. In other words, the change in the composition of these schools may have an effect on this group of pupils. Finally, the framework aims to capture the effect on the group of highly able pupils who would have attended grammar school even if the reform had not taken place and who may be affected by their relatively less able pupils.

The Northern Ireland data contain school-level data on the number of pupils (boys and girls) entering each year group and the School Leaver's Survey which consists of all school leavers from post-primary schools apart from pupils transferring to another school or those in special or independent schools. The data provides information on when students left school, the qualifications that they attained and their destinations after compulsory education. The data for England contain pupil-level information on A-levels as well as year group and gender (available from 1993 allowing consideration of the 1975 cohort outcomes). School-level information available from 1992 onwards and pupil-level information from 1993 onwards was used with regard to GCSEs. Mid-year estimates of cohort size are used for both Northern Ireland and England.

The authors found a marked discontinuity in the number of pupils entering grammar school around the time of the reform – the inflows increased by about 15 percentage points between the 1978 and 1979 birth cohorts (a 3.5 percentage point increase in the probability of attending grammar school) whereas it was reasonably stable for the four preceding and four subsequent cohorts.

They also found that this discontinuity was reflected in outcome measures: for example, there was an increase of about 12 percentage points in the number of pupils achieving one or more A-level and an increase of 17

percentage points for those achieving five or more GCSEs at A*-C at the time of the reform (between the 1978 and 1979 cohorts). Drawing a comparison with England, there was a higher probability of 2.4 percentage points of achieving 1+ A-level in Northern Ireland and a 4.6 point increase in the A-level measure of academic achievement. The authors, therefore, conclude that the reform must be the causal factor behind this difference as the two countries have 'exactly the same' examination system.

A comparison of cohorts born after and before 1978 by gender showed that the increase in the number of girls attending the more academic track was accompanied by an upward shift in their relative educational achievement. The probability of entering grammar school increased by almost 3 percentage points for girls during the 'open enrolment' period and this was also reflected in educational outcomes. The authors state that this provides evidence of the strong impact that grammar schools have on educational outcomes as, in this case, the process of entry into such schools has gender-related differential educational outcomes. As the authors put it, 'early differences between boys and girls contributes to later differences in educational achievement'.

They then conducted a similar experiment to explore whether the selective system exacerbates observed inequalities between socio-economic groups in relation to later educational outcomes. Comparing the effects of the reform on pupils coming from a poor family background and those coming from a more advantaged background (defined by their eligibility to receive free school meals), they found a large positive impact for both groups of pupils (an increase of +11% in the number of non-FSM students achieving one or more A-level and a +13% increase in the number of FSM students achieving the same qualification between the 1978 and 1979 cohorts). In other words, the reform had an equal impact on children with and without free school meals as far as grammar school entry and educational achievement at the ages of 16 and 18 was concerned. The authors concluded that 'an expansion of grammar school places is potentially beneficial to both groups but access to grammar schools is very unequal. Therefore, whatever pre-existing inequality there is between socio-economic groups (in terms of educational attainment at age 11) is exacerbated by the school system'.

The authors conclude that this 'open enrolment' reform has been followed by a clear impact in Northern Ireland relative to England: 'This suggests a strong causal effect by expanding the more academic track on overall educational attainment…this effect encompasses not only the direct effect of attending grammar school for the marginal entrants, but also the indirect effect arising through contextual impacts. The authors have mentioned this issue and the contextual situation but any possible impact has not been explicitly discussed.

We also note that the authors have not mentioned the number of schools and pupils that were considered in the research nor any specific reference is made to the school selection procedure. They claim that England has a 'fully comprehensive' system without mentioning at all its regional selectivity. According to Levačić and Marsch (2007), for instance, England still has around 20 local education authorities which are wholly or partially selective. The problematic nature of free school meals as an index of disadvantage needs also to be noted. These limit the extent to which it can be claimed that 'the academic track really has a causal impact…'.

3.2.15. *Clark (2007): Selection versus Comprehensives: Which Delivers the Best Educational Outcomes?*

In a recent paper published in the *Research in Public Policy* journal (2007a), Damon Clark, assistant professor of economics at the University of Florida, attempts to answer the question of whether children of different ability should be taught together or whether they should attend different types of school. He argues that credible comparisons across individuals, local authorities or countries that differ in their school systems are very difficult to be made (comparing, for instance, Britain's largely comprehensive system with Northern Ireland's selective system would hardly produce reliable conclusions because of important differences in their school systems). He states, however, that one type of within-authority comparison may produce credible results about the effects of attending grammar schools as opposed to attending secondary modern schools on borderline pupils. As Clark himself put it, 'the trade-off here is between an interesting question that we cannot credibly answer and a less interesting question that we can hope to answer convincingly'.

Clark made use of a dataset that contained information on several cohorts of pupils from the East Ridings (East Yorkshire) area who took the 11-plus in the early 1970s and attended twenty selective and non-selective schools. This dataset also included information on pupils' date of birth and gender, a measure of socio-economic status and the results of tests carried out in the fourth year of secondary school in addition to details of O-levels studied.

The author found that borderline pupils who attended grammar schools were likely to take more O-levels and, in particular, more advanced O-levels such as Latin and Greek. Selective schools, therefore, were found to have large effects on course-taking and ultimately on university enrolment, suggesting they may have important longer term impacts. Clark states that this finding is not surprising as secondary modern pupils do not have the opportunity to pursue these courses. Interestingly, this finding was particularly marked for pupils of lower socio-economic status who would be less likely to take O-levels in secondary modern schools. However, grammar schools were not found to improve performance on the fourth year maths test for those borderline pupils (Clark, 2007a). A four year attendance to a selective school had rather small effects on test scores (ordinary least squares estimates suggested effects of about one third of a standard deviation). Clark, therefore, reached the following conclusion: 'while grammar schools affected the number and type of O-levels pupils sat, they did not change basic learning outcomes, at least for pupils with borderline eleven-plus scores. …The East Ridings results suggest that grammar schools provided opportunities denied to pupils in secondary moderns, but did not improve basic learning outcomes'. In the full version of his paper, Clark (2007b) concluded that having high SES or highly-educated parents seemed to be of more benefit than attending a selective school considering the strong correlation between social class and parental education with test outcomes. In addition, the difference in peer quality between selective and non-selective school students was much larger than the test score effects.

On the basis of these results, Clark (2007a) concludes that 'the selective versus comprehensive issue is not important: grammar abolitionists are right to claim the transition to comprehensive education destroyed nothing of value; grammar supporters are right to claim the comprehensive movement

delivered nothing that could not have been achieved by widening secondary modern opportunities. The issue for historians and political scientists is whether these opportunities – the raising of the school leaving age, the merging of O-levels and CSEs into GCSEs and the expansion of further and higher education – could have been achieved without comprehensive schools' (p.17).

As Clark states, this study's results cannot be generalised as they are specific to one local authority and can only identify effects for borderline pupils (it is not possible to indicate how grammar schools can affect the most able pupils or how secondary modern schools can affect the least able ones). Further research using data from other authorities would be needed in order to assess the generalisability of these findings. In addition to this particular limitation of the study which restricts the analysis to only one authority, similarly to the NCDS data, the age of the data here makes it difficult to apply it to the current school situation.

3.3. *Updates on the NCDS data*

Here we report on more recent studies that have made use of the National Child Development Survey (NCDS) dataset. This is the dataset that was used by Steadman (1980, 1983) and is described in detail above. It is a dataset rich in information about respondents at various stages in their lives and as such is valuable to researchers as there are many explanatory variables that can be controlled for in analysing the data. However, from a secondary education perspective it must be noted that the respondents were age 16 in 1974, so some of this data is 30 years old or more, and was collected when the education system in England was somewhat different to what it is in 2005. So recent analysis of this data tells us more about different types of school in the 1970s than in 2005, but these recent investigators believe their results to still be of value.

3.3.1. Sullivan and Heath (2002): State and Private Schools in England and Wales

The primary aim of these authors was to investigate whether private education is superior to state education in terms of the examination performance of their pupils. Private schools were classified as independent or direct grant; state schools included grammar schools, secondary modern schools and comprehensive schools. After controlling for intake to the schools, the authors found that pupils at state grammar schools and private schools achieved superior educational outcomes to pupils at comprehensive schools.

Data

They used the National Child Development Study.

The authors note the age of this data and point out it cannot tell us about the current situation in Britain. However, the NCDS data is described by the authors as the richest British dataset currently available for exploring the questions they investigated. Their sample size was 10237 respondents (see Table 23 for the distribution of each type of school.)

Table 23: The distribution of the data by percentage at each type of school

	Direct Grant	Independent	Grammar	Secondary Modern	Comprehensive
Percent	2.4	4.3	12.6	25.0	55.8

Despite having individual and school-level data, they were not able to identify whether students attended the same school as other members of the sample. The sample was not clustered within schools, and it was very likely that many schools were represented by a single sample member. Thus there is no consideration of selective areas.

Explanatory variables

These were essentially the same as those used by Steedman, including a pupil's attainment at age 11, social class and family background and characteristics of the parents.

Outcome variables

The number of O' level passes or passes at CSE grade 1 and a mathematics test score at age 16. The reading comprehension test was the same as the one taken at age 11, so this was not used due to a suspected ceiling effect. The maths test was designed to be appropriate for the full ability range of 16 year olds.

Methodology

Multiple regression in which various models were investigated controlling for various explanatory variables.

The study was geared to answer these research questions:

1. How did the various types of school attended by the NCDS children vary in terms of student intake, in school resources and other school characteristics, and in students' test scores and examination performance?
2. To what extent are differences in academic outcomes for students at the different types of school explained by differences in the characteristics of the children who attended the schools? Are parents right to believe that, by paying fees for private education, they achieve better results for their children than they would obtained in a state school?
3. Are the better results, if any, obtained at private schools to be explained by the schools' financial resources, social capital or peer group processes? Social capital is described as social norms and networks, and said to exist in area surrounding a school where parents know each other and possibly the teachers.

Results and Conclusions

Pupil intake:

There were clear differences in the schools intakes of pupils. Private schools had privileged intakes in terms of pupil's cognitive skills and parents' social class, education, reading behaviour and interest in their child's education. The

academically selective direct grant schools had a lower proportion of parents from the employer and managerial classes but a greater proportion of the very able pupils than independent schools. The grammar schools' intake was similar to the direct grant schools but with a broader social class distribution. The comprehensive and secondary moderns schools had a relatively deprived intake in terms of both pupils' cognitive abilities and family characteristics.

Resources:

Independent schools had the lowest average pupil to teacher ratio (13:1), with the highest in secondary modern schools (18:1). In direct grant and grammar schools the ratio was about 16:1. However, the pupil to teacher ratio was found to have no effect on either test scores or examination results at age 16.

Attainment at age 16:

The authors limited their investigation into examination results to the top third of the ability range. However, these were said to largely agree with the results found for the tests results at age 16 which covered the full ability range. The test score at age 11 was found to be the most powerful predictor of educational success at age 16. Measures of cognitive skills at age 11 and of pupils' social backgrounds accounted for a substantial proportion of the difference between schools.

Social capital:

Using the measures available the authors found that home/school social capital did not account for differences in attainment outcome. They attributed this to the wide catchment areas usually associated with independent schools, and also with direct grant and grammar schools. They noted social capital is far more likely to influence outcomes in comprehensive schools that serve a more local community.

Social class composition of a school:

This was the only school level factor that appeared to explain some of the differences in outcome. The authors suggested this might be due to acceptability of norms regarding academic effort and success, and possibly through teacher expectations. It was noted that the effective implementation of homework implies a degree of teacher authority and discipline, which would be more prominent in the selective schools. The authors also speculated that there would be a greater emphasis on extra curricular activities such as sport, music and drama in the selective schools, and this might contribute to greater academic attainment at age 16.

Limitations

The authors note their attempt to control for individual pupil and family characteristics may not have been adequate, although what they did was said to be more thorough than in many other analyses of school effect.

The authors note the difficulty of allowing for parental choice. Parents can opt to pay the fees of an independent school and have a large degree of choice, whereas in the state system (1970s) the decision of which school a child attended was largely the decision of the local authority.

We note the biggest limitation is the age of the data and thus the inherent problems in interpreting the results for schools today. However, these authors raise several interesting points that are worthy of further research for present day schools, particularly the concept of social capital and peer group effect. These are related to school compositional effects which we discuss further in Chapter 8.

3.3.2. Galindo-Rueda and Vignoles (2004): The Heterogeneous Effect of Selection in Secondary Schools: Understanding the Changing Role of Ability

Galindo-Rueda and Vignoles (2004) made use of the NCDS data in their study. They explored the inter-relationship between school selection, ability, family background and educational achievement. Ability was taken to mean cognitive ability as measured by tests when the children were 7 years old. Data at age 7 was chosen in preference to data at age 11, as the latter may have been influenced by whether or not the pupil was subject to a selective system. They note the historical nature of the data in that these pupils were in secondary schools between 1969 and 1974, when comprehensive reorganisation was taking place across England to a greater or lesser extent in the LEAs. However they consider the degree of exposure of the children then to selective education and its effects on them, to be relevant to the debate on selection by ability today. Their principal finding was that the most able pupils in a selective system did better than those of similar ability in mixed ability school systems; the latter being the phrase these authors use for comprehensive schools. They found no significant effect due to selective systems on middle and low ability pupils. The measure of how well a pupil did in secondary school was essentially the result of a mathematics test at age 16.

The authors note, with the exception of Germany, several European countries have abandoned their selective systems of schooling from about 1950 onwards, yet selection remains a topical issue. They note recent government policy decisions in England have put pressure on schools to get better academic results. They note that implicit selection of pupils according to their perceived ability or parental background is a live issue particularly in systems called comprehensive, and that the issue of selection and how it interacts with family background and pupil ability is of great research interest.

Data

The National Child Development Study data set.

Sample size from 1974: 4715 pupils in comprehensive schools, 3198 pupils in either grammar or secondary modern schools.

Output variables

Score on a maths test at age 16; years of schooling; whether students went on to achieve A-level or higher qualification

The authors also used two indices on exposure to a selective system for use in their analyses. They defined these as:

1. Whether the pupil was in a grammar school or secondary modern school at age 16, or in a comprehensive school.

2. A measure of the number of years spent in a selective school system between the ages of 11 and 16.

We note that the latter is an interesting variant on other researchers who have used this data, as Galindo-Rueda and Vignoles have included those pupils who experienced both systems and the change over between them. Steedman (1983) separated the sample into two distinct groups of those pupils who had experienced one system or the other for all of the 5 years of secondary schooling, and rejected the pupils who had experienced a change of type of school from her analysis.

Explanatory variables

Ability at age 7 (this was subdivided into quintiles as Steadman had done); family background variables; characteristics of a child's neighbourhood; LEA educational resourcing level (further detail below).

Methodology

Multiple regression; matching technique

The matching technique was used to compare individuals with the same probability of being subject to selective schooling. The matching was based on area-level characteristics and educational characteristics at the level of an LEA.

Results and Conclusions

The most able pupils in the selective system did better than those who of similar ability in comprehensive schools, particularly the girls.

There was no significant effect of a selective system on pupils of middle and low ability.

Limitations

The non uniformity in the LEAs and the types of school within them was noted by the authors. They also noted the difficulty associated with a so called comprehensive school being close to a grammar school, which might attract the most able pupils.

We note all the explanatory variables available in the data illustrate the complexity of the problem and what factors might affect attainment and the likelihood of being part of a selective system. There is essentially only one output variable measuring attainment, a mathematics test at age 16. No other measures of attainment such as O -level or A' level results were considered nor was their any consideration of value added. Whether pupils went on to achieve A' levels or a higher qualification was considered, but numbers were very small and so not of significance. In the regression coefficient tables, standard errors were quoted but no effect sizes.

The age of this data and the systems it relates to, limit its applicability to modern day. The authors noted in their introduction how the curriculum at grammar schools differed very much from that at secondary modern schools. Also secondary modern pupils took CSE examinations while grammar school pupils took the more academically demanding O-level examinations. We note this is of little relevance some 30 years later with the advent of The National Curriculum and GCSE.

Further detail

An initial analysis of the data indicated that the more able children from wealthy backgrounds were more likely to be educated in grammar schools.

It was also noted that research by others indicated that LEAs that had been under Conservative control changed more slowly towards comprehensive reorganisation than those under Labour control. In the present analysis the authors found pupils were more likely to be in comprehensive schools in areas under Labour control.

Explanatory variables:

> Family background: father's social class, number of years parents were in school, details of siblings and household income.
>
> Area level characteristics: included census-based characteristics at the level of an individual's enumeration district – reflecting the socio-economic group distribution, proportion of immigrants and proportions doing professional, semi skilled or manual jobs and the political affiliation of the child's constituency.
>
> LEA resourcing level: teacher – pupil ratios, costs per student, number of pupils per 1000 head of population.

All these variables, together with the measure of ability at age 7, were chosen to attempt to control for the complexity inherent in the degree of change over from selective to comprehensive systems in different areas and thus to bring out the effect of exposure of a child to a selective system.

Results of regression with output variable number of years in a grammar or secondary modern school (i.e. in a selective system).

It was noted that most of the explanatory variables were not of any significance, but the political affinity of the child's constituency did indicate that children in a Conservative constituency experienced on average 0.7 years more years of selective schooling. Children from the top quintile of ability also spent more time in a selective system, particularly girls. (girls 0.4 years and boys 0.1 years) (We note this is only of historic interest but does support the view that Conservative areas were slower to change their secondary school systems).

Results of matching considering educational outcomes.

The educational outcomes considered were mathematics and reading scores at ages 11 and 16, a general ability score at age 11, number of years of schooling and whether the pupil went gained A' level or higher education qualifications. Of these scores at age 11 were considered dubious due to the possibility of coaching for the 11+ and distortion of the primary school curriculum. The reading test was the same at age 11 as at 16 and so regarded as being of limited value. The results with most impact were for the highest quintile, particularly for girls.

The authors further analysed the above results so that grammar schools could be considered separately to secondary modern schools rather than the two together forming a selective system.

Table 24 below is interesting as it is indicative of what happens if a pupil is "misclassified" at age 11. Looking at the top quintile, it can be seen that 189

pupils of high ability (as measured at age 7) went to a secondary modern school and suffered adverse affects as a result as evidenced by their negative mathematics scores. It can also be seen that from the lowest quintile 11 pupils went onto grammar school and benefited greatly from doing so.

Table 24: Matching estimates of the effects of selective schooling, decomposition by ability quintiles and selective school type. Outcome variable: maths scores at 16. Extracted from Galindo-Rueda and Vignoles (2004)

	Boys low high			girls low high		
	Abil Q1	Abil Q3	Abil Q5	Abil Q1	Abil Q3	Abil Q5
Selective vs comprehensive	0.973	1.189	1.916	-0.203	1.328	3.016
Secondary modern vs comprehensive In ( ) no. of pupils in secondary modern	0.627 (215)	-0.424 (181)	-4.052 (89)	-0.537 (184)	-1.018 (189)	-3.379 (90)
Grammar vs comprehensive In ( ) no. of pupils in grammar schools	13.800 (5)	6.638 (51)	4.403 (200)	9.166 (6)	7.205 (70)	5.066 (265)

The authors note that the negative effect on those high ability children who failed to get into a grammar school was an argument used by many people in favour of comprehensive education. We note it can be argued both ways, that those of apparent low ability make outstanding progress once in a grammar school. We also note the limitations of the outcome measures, but these results do illustrate the central questions of what is ability and is it static, how are decisions re selection made at age 11 and how does this affect outcome at age 16.

Results of regression with output variables as mathematics score at 16, years of schooling, A-level or higher education qualification, shown separately for the two indicators, selective v non-selective at age 16 and years of pre 16 selective schooling. The results again show the more able benefited from attending a grammar school. It is notable that the coefficients given for years of pre-16 schooling are about a quarter of the size of those for selective v non selective schooling although the authors note the relative sizes are the same. The authors conclude:

> Children, especially girls, in the top of the ability distribution did attain better educational outcomes if they were in a selective school system.

In discussing their results, the authors cite another of their papers, which indicated that early ability has started to play a lesser role in determining how well someone does at school, whilst family background is growing in importance. They take this further in the present study by saying the move to comprehensive schoolings has disproportionately benefited the less able, but wealthier, students and the gap (in educational attainment) between the most and least able has been reduced. They again noted how their results indicated for the most able (top quintile or 20%) that these pupils did do better in a selective school system, particularly the girls.

The authors conclude their paper with the speculation that the housing market might have taken the place of the 11+, in granting parents the right to

send their children to schools where demand exceeds the supply of places; they noted many of these schools are former grammar schools.

3.3.3. *Manning and Pischke 2004: Ability Tracking and Student Performance in Secondary Schools in England and Wales*

It is interesting that a second pair of education economists from the LSE chose to analyse the NCDS data set despite the limitations due to the age of this data and the children it refers to being in secondary schools in the period 1969 to 1974. We will not review this research in the same depth as previous studies, not least because these authors stress the tentative nature of their findings. We have also given extensive discussion and comment on Steedman's original analysis of this data (1981, 1983) and the more recent updates. In particular we have noted the relatively small sample and the large degree of attrition suffered by this data set as the children grew older. The data set analysed by these authors had 11,407 individuals left in the study, this being when they were 33 years old.

These authors note that much previous research into selectivity has focussed on the effect of attending a selective school on the performance of the individual concerned. They focus on the question "what is the affect of the availability of selective schools in an area on the performance of pupils, irrespective of whether they actually attend a selective school". They found that selective schools tend to perform at least as well or better overall, but there may be an advantage for some children from attending comprehensive schools, particularly those with high ability but from a poor background.

It could be argued that any results from analysing this data are now only of historical interest. One reason for the continued interest in the NCDS data maybe because it is very rich in socio-economic data, and this is lacking in the up to date national pupil data sets collected by the DfES. However, the NCDS data is lacking in a variety of performance measures, and it is notable that these authors only used the mathematics test score at age 16 as the performance outcome variable in their regression analyses.

The authors took the view that it is necessary to compare pupils in comprehensive LEAs to pupils in selective areas rather than comparing pupils in comprehensive schools to pupils in selective schools to pursue their question about the affect of the availability of selective schools in an area. However, they encountered the problem of defining such LEAs as we have highlighted this as problematic in our review of other studies. This problem is illustrated by the definition that these authors reached for the two types of LEA after investigating various possibilities.

They said that for a LEA to be comprehensive it required 75% of its pupils to be in comprehensive schools in 1971 and 1974, and more than 10% in grammar schools in 1967 but less than 5% in 1974.

They said that for a LEA to be selective it required to have less than 20% of its pupils in comprehensive schools in 1971 and less than 40% in 1974. In addition they required the LEA in 1974 to have at least 10% of its pupils in grammar schools and that the fraction in grammar schools in 1974 should be at least 80% of the 1967 fraction.

In 1974, according to these definitions, about 12% of pupils age 16 in a selective LEA attended a comprehensive school, and over 90% of pupils age 16 in a comprehensive LEA.

In their analysis the authors also included data on the political affiliations in the areas concerned using local election results from 1961 and 1967, but this appeared to have no affect on test performance outcome.

The authors found that comprehensive LEAs tended to be poorer than selective LEAs as evidenced by the socio-economic status of the parents and the poorer housing stock. They also found student performance to be worse in comprehensive LEAs according to both teacher assessment and to formal tests. They also noted that pupils living in a comprehensive LEA were less likely to attend an independent school.

The authors note that pupils from the comprehensive LEAs passed significantly fewer O-level and A-level examinations than those from selective areas, but that ultimately they earned higher wages. They stated that their overall results indicate some negative effects on learning in comprehensive schools, but that any detrimental impact of comprehensive schooling might be short lived. The authors say this could possibly be attributed to the nature of comprehensive schools at the time; the pupils might have been disrupted by reorganisation and that comprehensive schools tended to be bigger than selective schools and to have bigger classes. They found that pupils in comprehensive schools liked their schools less than those from selective LEAs.

In terms of particular types of pupil, the authors found that high ability children from a high socio-economic background benefit from a selective system, whereas high ability children from a low socio-economic background might actually be better of in a comprehensive school.

We note again that these results are an interesting reflection on the nature of selective and comprehensive systems in the reorganisation era of 1969 to 1974, but have little relevance for today other than suggesting further research using up to date data.

3.3.4. *Manning, A. & Pischke, J. (2006). Comprehensive versus Selective Schooling in England and Wales: What Do We Know?*

In this paper, Manning and Pischke attempt to demonstrate that studies using value-added measures are unlikely to be entirely successful in eliminating selection bias between comprehensive and selective school students. They again use the National Child Development Study as their main data source and analyse the maths test score at age 16 as the key outcome measure. They argue that the measure of exposure to a comprehensive or selective system is unclear. In addition, they claim that comprehensive areas were systematically poorer and had students with lower prior achievement than selective areas.

The authors propose what they call 'a falsification test for the value added specification' where value added models are applied on student performance at age 11 as the dependent variable, and controlling for student performance at age 7. They argue that, as 11 year old students have not started attending secondary school yet, performance outcomes should not be affected by the secondary school environment if all selection is successfully controlled, in other words, if the selection bias is successfully removed.

However, they have found similar patterns of results for the age 11 and 16 test scores. They conclude that age 11 results may reflect selection bias. As this may also be the case for the age 16 test scores, they argue that caution is required in interpreting the age 16 results causally, i.e. that any differential

outcomes in performance are due to the comprehensive or selective schooling system.

To account for the possibility that their age 11 results are biased (on the basis of arguments such as that age 7 student performance is a poor predictor of age 11 outcomes or that there may be variation in primary school teaching depending on whether pupils plan to take the 11+ exam), they attempted to control for a variety of factors which showed that these are unlikely to explain the age 11 results.

They carried out OLS regression analyses using, firstly, the math test score at 16 as the dependent variable and test scores at age 11 as the controls and, secondly, the math test score at 11 as the dependent variable and age 7 test scores as the controls. Using the technique of adding additional covariates, for instance, other age 11 test scores (a reading, verbal, non-verbal and a design copy score), demographic and family background variables (similar to the ones use in Kerkhoff *et al*, 1996), they argue that, if the strategy worked well in controlling for all selection, the comprehensive school coefficient when using the test score at 11 as the dependent variable would be expected to be zero. However, its absolute value is twice as large as the coefficient in the case of math test scores at 16, possibly reflecting selection bias in both cases. The authors accept though the possibility of the primary school experience in spending time getting prepared for the 11+ examination having an impact on student achievement although they doubt that it could have twice as large an impact than actually attending a comprehensive school.

A further analysis aimed to explore whether the larger degree of measurement error in age 7 test scores could partly explain the larger differences found for age 11 math test scores. It was found that even though measurement error is likely to lead to bias, it does not seem to be the main explanation as strong negative effects for age 11 scores are still found. The authors conclude that 'selection remains a plausible and maybe even likely explanation for these findings, and hence sheds doubt on the age 16 results as well'.

More analyses on 1. interaction effects between attending a comprehensive school and ability, 2. using the variation at the LEA level by comparing wholly selective or whole comprehensive LEAs and 3. using political control of the county as an instrument for early implementation of the comprehensive status (instrumental variables strategy) do not seem to be successful in removing the selection bias.

The authors conclude that it is *a possibility* (their emphasis) that the age 16 results are right and the age 11 results are wrong. Their evidence indicates that 'there is a good case to be made that selection bias exists in the estimates comparing students in comprehensive and selective schools… we probably do not know very much about the effect of comprehensive schooling in Britain, or elsewhere for that matter'. In spite of the valuable and rich information that the NCDS dataset provides, the age of the data limits its applicability to the current school situation as has been mentioned previously.

3.4. Other countries

3.4.1. Research from Northern Ireland

A research commission funded by the Northern Ireland Department of Education in 1997 executed an extensive investigation into the effects of the Northern Ireland selective system of secondary and grammar schools. Researchers from Queens University Belfast, the University of Ulster and other research organisations carried out the research. The research project included various kinds of data gathered from 8 grammar and 17 secondary schools across Northern Ireland. The detailed evidence was discussed in a series of research papers which arose from an extensive body of fieldwork in schools in Northern Ireland and Scotland. The purpose of the research was to provide informed basis for discussion on the future of the selection system in Northern Ireland.

The data for the research project were gained via interviews with teachers and pupils in secondary and grammar schools, focus group interviews with certain groups, observations, postal questionnaires as well as statistical data held by several sources. Analyses were done to look at the impact of selection on primary schools, post-primary schools, teachers, pupils and society. For the purpose of the present study, only studies related to the impact of selection on pupils are being covered. The evidence collected was related to GCSE achievement, impact of selection on pupils' post-school destinations and impact on their motivation and attitudes.

Achievement at GCSE

Shuttleworth and Daly (2000) did a study looking at achievement at GCSE by focusing on GCSE results using school-level and pupil-level data. School-level data revealed that there is a clear difference in the achievement levels of grammar and secondary schools, with the average levels of achievement being higher among grammar schools. On average, the proportion of year 12 pupils passing 5 or more GCSEs at grade A* to C in 1997/98 was 95% for grammar schools and 31% for secondary schools. The achievement level of grammar schools was uniformly high but there was a high degree of variability among secondary schools. The inter-quartile (difference between schools at the 25th and 75th percentile) range for grammar schools was 5% but was 18% for secondary schools.

The pupils-level data were collected from 1,784 Year 12 pupils in a sample of 8 grammar and 17 secondary schools (academic year 1998/99). Information was sought on individual and family background, and attitudes to schooling via questionnaire. Analysis of pupil data for Transfer Grade showed that practically all A grade pupils and the majority of B grade pupils were in grammar schools, while the majority of C grade and practically all of D grade pupils were in secondary schools. The Transfer Procedure at age 11 involves an 11+ test in which the top 25% of children are warded grade A, the next 10% are awarded grade B, the next 10% are awarded grade C and the rest are awarded grade D.

Analysis of the pupil-level data showed strong evidence of a 'grammar school' effect. Results via multi-level regression indicated that grammar school attendance had a major positive effect on GCSE scores which was independent of, and additional to the influence of the other explanatory

variables included in the model (transfer grade, gender, entitlement for free school meals, father's occupation and number of siblings). The results showed that being in a grammar school added 15.95 GCSE points, equivalent to three GCSEs at grade C, to a pupil's achievement at age 16. Nevertheless, this study did not include analyses to investigate who are the ones actually benefiting from the grammar schools (e.g. borderline pupils as in Schagen & Schagen, 2003).

Selection and post-school destination

Daly and Shuttleworth (2000), in a separate analysis, examined the extent to which there were differences in the post-compulsory education experiences of young people in the secondary and grammar sectors, and to describe the various individual, family and school factors which were related to differing post-compulsory education statuses.

Leavers from grammar and secondary-school sectors, on average, follow very different 'tracks' after compulsory education is completed. The vast majority of grammar school pupils remain at the same school until the age of 18 (80.9%), but only a small percentage of secondary school pupils did so (15.8%). The rest of the secondary school pupils entered a wider range of destinations including other school (6.6%), training (19.5%), employment (10%) and further education (47% to FE college).

Multi-level logistic regression analysis was employed to examine the extent to which type of school attended (grammar/ secondary) had an independent effect on post-compulsory education behaviour once other variables had been taken into account. Results show that even when transfer grade, gender, entitlement to FSM, father's occupation and number of siblings were taken into account, grammar school had a positive and independent effect on the chances of staying at the same school, but was not a statistically significant predictor of whether a pupil remained in post-compulsory education. However, it is possible that grammar school pupils are more likely to stay in the same school because grammar school pupils are more likely to achieve better and take A-levels which are normally offered at their schools.

Selection and pupil motivation and social attitudes

Gallagher and McKeown (2000) examined the attitudes of pupils in their schools. The data were collected from 2,130 Year 12 pupils in the schools used as case studies in the project. Attitudinal data were collected from Year 12 pupils in a sample of grammar and secondary schools, and from a sample of schools in Scotland. A questionnaire developed by the Australian Council for Educational Research was used by the researchers (ACER School Life questionnaire). The questionnaire included a series of measures that tapped positive and negative attitudes towards school, pupils' sense of social integrity, status and success, and attitudes to teachers and to the perceived relevance of the curriculum. The study included a focus on differences due to school type.

The results showed that there were statistically significant differences between grammar and secondary school pupils in negative affect, teacher, relevance, success and school integration scales (table below). However, there were no statistically significant differences between the schools in terms of general satisfaction and status.

Table 25: Mean, standard deviation and F values for ACER School Life variables by school type.

ACER scales	Grammar (n=949)	Secondary (n=1,180)	Significant effect of school type
Negative affect	2.03 (0.50)	1.98 (0.51)	F= 6.8; p<0.001
Teacher	2.86 (0.47)	2.96 (0.50)	F=23.3, p<0.000
Relevance	3.10 (0.46)	3.15 (0.45)	F=4.2, p<0.05
Success	2.98 (0.42)	2.93 (0.45)	F=5.9, p<0.01
School integration	3.00 (0.37)	3.04 (0.43)	F=5.0, p<0.05

( ) standard deviation

The paper suggests that grammar school pupils were aware of the high academic standards expected of them and many of them felt under pressure because of the high expectations. The authors defined this as evidence that the pupils experienced a little less enjoyment in some aspects of their school experience in comparison with their peers in secondary schools. At the same time however, the grammar school pupils had a high expectation that they would succeed in school. Grammar school pupils placed a high priority on the achievement of high academic results, while the secondary school pupils placed high priority on the provision of a supportive and caring environment. Although there were some differences in the expressed views of grammar and secondary school pupils, all pupils appeared to be positively disposed towards their schools.

It seems that the different types of schools serve different purposes for different types of pupils, and in general it didn't matter to the secondary modern school children that they were in secondary modern schools as it's the supportive and caring environment that they were looking for. As for the children of higher ability (in grammar schools), they were obviously anxious due to the high demanding environment, being around peers of high ability, but that environment is probably the best environment to realise their true potentials (although they were anxious, they were still confident that they would be successful).

3.4.2. Croxford (2000): Inequality in Attainment at 16: A 'Home International' Comparison

Croxford (2000) investigated whether the differences between the four UK education systems (England, Northern Ireland, Wales and Scotland) are associated with different levels of social inequality in attainment. The project reviewed existing statistics and research, interviewed policy-makers in the four territories, and constructed and analysed an integrated dataset in the early 1990s. This dataset was based on the England and Wales Youth Cohort Study of young people aged 16 on August 31, 1990, surveyed in spring 1991, spring 1992 and spring 1993, The Scottish Young People's Survey of the year group in S4 in 1989-90, surveyed in spring 1991 and autumn 1993, and the Northern Ireland Secondary Education Leavers' Survey of secondary school leavers, surveyed in 1992 and 1994/5. The study examined differences between the four education systems in the pupil and school characteristics which influenced attainment.

In spite of the different definitions of the national curriculum in each system, the curriculum between the four systems is broadly similar. Pupils study seven or eight general subject courses during the last two years of compulsory schooling, and public examinations at around 16 provide

certification of the pupils' attainments during their time at school. In Scotland, the examinations are known as the Scottish Certificate of Education (SCE), while in the other three systems, they are known as the General Certificate of Secondary Education (GCSE).

The greatest difference between the four systems is the pattern of comprehensive versus selective schooling. Amongst the pupils attending state schools in 1990/91, all pupils in Scotland and almost all pupils in Wales were educated in comprehensive schools. In Northern Ireland, the system is entirely selective. Regional differences existed in England in the extent of selection, and only a minority (7%) of pupils were educated in selective schools. Beside selection into grammar schools, many potential pupils were also 'creamed' to the private sectors. The proportion of pupils attending independent schools also varied (England: 11%, Wales: 3%, Scotland: 6%).

Differences also occurred in terms of the proportion of pupils attending denominational schools (England: 19%, Wales: 7%, Scotland: 14%, Northern Ireland: 100%), and single-sex schools (Scotland: 3%, Wales: 15%, England: 18% and Northern Ireland: 33%). The study also found that there was more social segregation between schools in England and Northern Ireland than in Wales and Scotland. In England and Northern Ireland, pupils attending grammar and independent schools tended to come from higher social background (SES) than pupils attending comprehensive schools or secondary modern schools. In Scotland, there were more pupils with relatively high SES attended comprehensive schools than was the case elsewhere. Nevertheless, differences also existed in the social composition between comprehensive schools in terms of the average SES of pupils attending the schools. This was due to differences between catchment areas and may be intensified by parental choice of schools.

Results showed that on average, attainment was higher in independent and grammar schools, followed by comprehensive schools, and lowest in secondary modern schools. Attainment was also higher in Roman Catholic schools and single-sex schools than in non-denominational or mixed-sex schools. These differences were found to be similar in all areas with each type of school, but the advantages of independent schools and Roman Catholic schools were weaker in Scotland than elsewhere.

Social composition was found to have an effect on pupils' attainment. Schools with higher average SES had higher average attainment than schools with relatively lower average SES. The effect of school social composition was weaker in Scotland, Wales and Northern Ireland than in England, irrespective whether the school was selective or not.

There was higher variability in attainment between schools in England and Northern Ireland after taking account of all other factors. Schools in Scotland and Wales were much more similar in their average attainment. Croxford could not find any evidence that comprehensive schooling depressed the overall levels of attainment as in Wales, which had a fully comprehensive system and very few independent schools, the average attainment was no different from that of England and Northern Ireland. Scotland could be used as comparison due to the different examination system.

Croxford concluded that the evidence from this study provided support for the comprehensive system. Several points were highlighted in supporting her conclusion:

1. there was less social-class inequality in the two education systems which are wholly comprehensive;
2. selection was not the only source of inequality within education systems. Other sources of inequality include independent schools, denominational schools, single-sex schools and social segregation between schools;
3. there was no evidence that the overall levels of attainment were affected by the extent of selective or comprehensive schooling;
4. standards across schools were more uniform in the two systems which have least differentiation between schools.

This study highlights several sources of inequality within the education system. Besides the selective and comprehensive systems, other sources of inequality include independent schools, denominational schools, single-sex schools and social segregation between schools. The study provided support for the comprehensive system, as the effects of the other inequalities were less in the comprehensive education system, as in Scotland and Wales. Figures and results of statistical analysis were not presented in detail in the paper, therefore further comments could not be made.

3.4.3. Croxford & Paterson (2006): Trends in social class segregation between schools in England, Wales and Scotland since 1984

In this paper, Croxford and Paterson aimed to describe overall trends in social class segregation between British secondary schools, and to particularly explore whether there was a difference in those trends in the more comprehensive school systems in Scotland and Wales compared to those in England. The authors aimed to address the following research questions:

1. To what extent is there segregation between schools of students from different social class backgrounds? Does the extent of segregation differ in England, Wales and Scotland? If there are differences in segregation, do these support the hypothesis that segregation is lower in more comprehensive education systems?
2. Has the extent of segregation changed during the 1980s and 1990s? If levels of segregation have changed, do they support the hypothesis that parental choice legislation has increased segregation? Is there evidence of a 'starting-gun' effect and subsequent decline?
3. Have there been different patterns of change in England, Wales and Scotland? If there are differences, do they suggest that parental choice has had different effects in each national system?

The analysis is based on the England and Wales Youth Cohort Study (YCS) and the Scottish School Leavers' Survey (SSLS) both comprising young people in the last year of compulsory schooling. Those datasets are comparable as their purpose and content are similar and both have been carried out on a regular basis since 1985. The SSLS samples consist of all secondary schools including independent schools but excluding special schools. The types of schools attended by the YCS samples included independent, grammar, modern and comprehensive schools as well as city technology colleges. Selection for the YCS cohort samples was more complex and problematic than for SSLS, and the authors claim that some relevant changes that had to be adopted had implications for the subsequent analysis. Response rates to

both surveys have declined over time, the average fraction of the population of 16 year old students in the relevant year being 2% in England and Wales and 8% in Scotland.

The measures of socio-economic status used for the segregation indices were derived from parents' occupational status and education. These were:

- Working class family (vs. other two classes, e.g. managerial and professional, and intermediate)
- Managerial and professional family (vs. other two classes, i.e. intermediate and working class)
- Composite SES measure (this variable was not calculated for England and Wales cohorts for 1984-1988 as relevant data on parental occupation and education were missing)

The authors note the study's limitations. Even though the Scottish part of the study has been consistent in design, sampling and coding, the authors mention 'its major gaps' regarding the early 1990s cohorts. The England and Wales YCS parts of the study are less reliable due to their inconsistencies regarding, in particular, the coding of parents' occupation and sampling procedures which had to change significantly. The authors also had limited information available on placing requests indicating parental choice exercised by survey respondents. However, the strength of the analysis lies, according to the authors, on 'the length of time over which the data are available, the opportunity to compare the three British education systems, and the scope for studying segregation of social classes (defined by parental occupation) rather than by administrative variables such as the proportion entitled to free school meals'. Addressing their three research questions, the authors reached the following conclusions:

Some differences have been found in levels of segregation between the national systems but these have not been as large as would have been expected. Segregation for working class pupils was consistently lower in Scotland than in England, a finding which suggests that lower segregation resulted from the more comprehensive system in Scotland. There were no differences, however, in the pattern of segregation for managerial and professional class pupils in the three countries. Some evidence found that segregation was lower in Wales for the 1980s cohorts and particularly for managerial and professional class pupils up to 1993, may also lend support to the hypothesis that segregation is lower in more comprehensive education systems. However, these findings are rather tentative as they are limited by small sample size and changing sampling methods.

In relation to segregation trends between 1984 and 1999 and whether any differential trends found in the three national systems are associated with parental choice, the following findings were reported. No consistent increase in the level of segregation in the 1980s and 1990s was found. Some trends were clearer for Scotland than England and Wales: some evidence indicated that working class pupils became more unevenly spread across schools (indicating this group's over-representation in some schools or under-representation in others), and that managerial and professional pupils became more isolated (indicating this group's isolation from the mainstream, i.e. the majority group); working class pupils, however, became less isolated. Some evidence of a rise in segregation in England and Wales after the 1988 Education Reform Act which was followed by decline, lends support to the

'starting-gun effect'. The authors, however, point out the limitations of their study in rendering any more conclusions speculative. They maintain though that, as their findings suggest that 'the more comprehensive system in Scotland is associated with lower segregation and the more diversified system in England with greater segregation', they have future policy implications especially as the current education policy in England is encouraging greater school diversity.

3.4.4. Marsh (1991): The failure of high-ability high schools to deliver academic benefits: the importance of academic self-concept and educational aspirations

In Australia, Marsh (1991) explored the effects of the selective system onto pupils' academic self-concept as well as their educational and occupational aspirations in his study. Research emphasizing a psychological perspective of social comparison processes shows that school-average ability (SAA) is negatively associated with academic self-concepts (ASC). Sociological research indicates that SAA is negatively related to educational and occupational aspirations. Marsh (1991) united these two related research areas, and extends the diversity of outcomes and the theoretical frameworks considered. He used a longitudinal data of High School and Beyond (HSB) in which the effect of SAA on a comprehensive set of academic outcomes (eg, standardized test scores, ASC, course-work selection, academic effort, school grades, educational and occupational aspirations, and college attendance) was measured in the sophomore and senior years of high school and 2 years after high school graduation in Australia. The subjects included 10,613 respondents selected for the second follow-up of the sophomore cohort of the HSB study; all attended the same high school as sophomores and seniors. For the purposes of statistical testing, a sample size of 4,000 was used.

Path analysis was used to test relations among 23 variables: sex, socio-economic status, academic ability, school-average ability, school-average-SES, variables measured at sophomore year (T1): general self-concept, academic self-concept, coursework, effort, grade point average, educational aspirations, occupational aspirations, variables measured at senior year (T2): ability, general self-concept, academic self-concept, coursework, Effort, Grade Point Average, Educational aspirations, occupational aspirations, and variables measured at post secondary (T3): college attendance, educational aspiration and occupational aspirations. Variables at T1, T2 and T3 were outcome variables. Path model was based on the temporal order of the variables, Bandura's theory of social cognition, 'Big Fish Little Pond Effect', school context research and motivational theories.

The influence of school-average ability (SAA) was not positive for any of the outcomes at any time and was moderately negative for some. SAA most negatively influenced academic self-concept (betas between -0.20 and -0.23), and had somewhat smaller negative association with course selection, GPA, occupational aspirations, and to a lesser extent general self-concept and college attendance (betas between -0.08 to -0.18). SAA's relationship with subsequent standardised test scores (T2) was very small (beta = -0.03), and SAA's effect on effort was not statistically significant. The negative SAA effects persisted even after controlling for T1 outcomes. Many of the negative effects of SAA were found to be mediated by the combination of academic self-concept (T1 and T2) and educational aspirations (T1).

Marsh concluded that the academic outcomes related to higher-ability schools were not commensurate with the ability levels of students attending these schools, and no academic advantages of such schools were observed for any outcomes. Marsh stated that even though the disadvantages of attending higher-ability schools may not generalise to all higher-ability schools and to all individual students, the results demonstrate that it is unjustified to assume that attending higher-ability schools will necessarily result in any academic advantages. On the basis of this study and previous research, it appears that higher-ability schools do not provide academic benefits beyond those provided by lower-ability schools and apparently disadvantage at least some students attending these schools. Marsh suggested that the BFLPE effect could be altered by changing the competitive orientation of the school and the nature of feedback provided to students. However, it is important to point out that Marsh's study was based on Australian sample. The education system in Australia may be different from the education system in England. For example, the nature of assessment is Australia is mainly based on teacher assessment and therefore educational attainment is subjective to teacher variance. It is possible that teachers in high ability schools are more demanding and have higher expectations from their pupils as they are used to highly able pupils.

The increasing availability of data with internationally comparable test scores has recently generated research on the effect of tracking on students' test scores. Two recent papers (Hanushek & Wößmann, 2006 and Waldinger, 2007) have looked at whether ability tracking and selection across countries leads to inequality in educational outcomes. These papers apply a differences-in-differences approach which allows them to compare outcomes between secondary and primary schooling. Hanushek & Wößmann conclude that tracking raises educational inequality in contrast to Waldinger who does not find such evidence. However, Waldinger argues that the results provided by Hanushek and Wößmann are not robust to alternative sample or tracking measures. Jenkins, Micklewright and Schnepf (2006) looked at levels of social segregation drawing comparisons between English secondary schools and the situation in other countries.

These studies are discussed in more detail below.

3.4.5. *Hanushek & Wößmann (2006): Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries*

Hanushek and Wößmann attempt to avoid the difficult selection problems which arise when trying to evaluate the effects of 'tracking' within particular countries as they maintain that it is difficult to separate its impact from other influences on achievement. They aim, therefore, to address the problem of unobserved country level variables by adopting a differences-in-differences strategy.

They use international data sets (PISA, TIMSS and PIRLS datasets testing reading, mathematics and science) administered to students in secondary and primary school. They identify tracking effects by comparing performance differences between primary and secondary schools across tracked and non-tracked systems using the 'macro variation' in the institutional structure of between-school tracking and student performance in different countries. Their analysis, in other words, uses cross-country variation in tracking to

identify its effect on the within-country variance of educational test scores. Each country's primary school outcome is used as a control for its secondary school outcome. Their differences-in-differences estimates allow them, for instance, to measure the extent of educational inequality which may be present in late primary school before actual tracking takes place.

The authors look at whether tracking increases inequality measured by three different measures of inequality: the within country standard deviation, the test score difference between the student performing at the 75th percentile and the student performing at the 25th percentile in each country, and the 95th-5th percentile differences of test scores. Their analysis indicates that early tracking increases inequality in achievement. They find, however, mixed evidence about possible efficiency gains from tracking. The results on reading and mathematics, for instance, indicate a statistically significant lower achievement associated with early tracking whereas for science two of the three estimates indicate positive achievement effects from early tracking.

They argue that their analysis 'provides reasonably strong support for the unequalising effects of early tracking' and argue that these preliminary results suggest that 'countries lose in terms of the distribution of outcomes, and possibly also in levels of outcomes, by pursuing such policies'. However, Waldinger (2007), as discussed in the following section, challenged their findings showing that these are not stable to using a different measure for the tracking regime and to restricting the sample to OECD countries (Organisation for Economic Cooperation and Development). Waldinger also states that an important problem of their study is the 'extremely small sample sizes' which result from the use of 'country-level data of the dispersion of test scores'.

3.4.6. Waldinger, F. (2007): Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores?

Waldinger is interested in whether the importance of family background is stronger or weaker in countries that differ in the extent to which they track pupils at an early age. He challenges the interpretation of a positive correlation that some studies have found between tracking and the importance of the family background for students' test scores as a causal effect. Using a difference-in-differences methodology, he concludes that tracking (defined in this paper as educating students in different types of schools rather than tracking for certain subjects within a school) does not exacerbate the importance of family background for educational achievement once pre-tracking differences are controlled for. Waldinger argues that unobserved country-level factors may exist and intensify the importance of parental background on students' test scores.

The author uses cross-country differences in tracking policies by making use of data from three large international educational studies: PISA (Programme for International Student Assessment), PIRLS (Progress in International Reading Literacy Study), and TIMSS (Third International Mathematics and Science Study). Data are used on test scores (reading and mathematics), student characteristics, parental background and school quality variables with a country level measure of the tracking regime indicating early or late tracking. Only data from OECD countries are used to compare countries with a similar educational development. Using tests administered to students before tracking has taken place in any of the countries in the sample (PIRLS

testing students in grade 4 and TIMSS mathematics data from grades 3 and 4), the author finds a negative relationship between the importance of parental background and the tracking regime suggesting the existence of unobserved factors which are correlated with a country's tracking regime and affect the impact of the family background.

Waldinger attempts, therefore, to control for these pre-existing differences across countries that implement tracking at an early or later stage. He concludes that his difference-in-differences results show that 'once the "pre-tracking" level of the family background effect on children's test scores is controlled for, tracking no longer affects the impact of family background'. He states that studies looking carefully at the effects of tracking are important in order to understand its effect on educational inequality and inform policy making in addition to educational research.

The author's difference-in-differences strategy looks at test scores taken at two points of a child's education, one in primary school before tracking has taken place in any of the countries considered in the sample and the second in secondary school after tracking has been implemented in some countries. The change between the two tests in the importance of family background in early and late tracking countries is then compared.

Results

Regarding the reading results, Waldinger reports the following findings:

- The importance of parental education for reading is as important for primary school students as it is for secondary school students. Parental education is found to be more important in early tracking countries. However, the analysis suggests that parental education does not become more important after actual tracking has taken place.
- The number of books in a student's home used as the relevant family background variable, is more important for secondary school students than for primary school students, and it is more important in early tracking countries. However, after actual tracking has taken place, the number of books in a student's home do not become more important in early tracking countries.
- Speaking the test language at home is a significant factor of doing well in reading and it is more important in primary school than it is in secondary school. It is also more important in early tracking countries. Again, this factor does not become more important after tracking has taken place.

Waldinger concludes that 'the results cast serious doubt on interpreting the correlation of tracking and the importance of the family background as causal'.

In relation to mathematics, the analysis reveals the following:

- The factor of the number of books in a student's home is an important one affecting student performance particularly in later stages of a student's education and becomes less important in early tracking countries after actual tracking has taken place.
- Speaking the test language at home significantly affects mathematics test scores and seems to be more important in primary school. It

seems to be more important in early tracking countries but it does not become more important after tracking has taken place.

Waldinger again concludes that the results do not lend support to the hypothesis that tracking exacerbates the role of family background for students' test scores after actual tracking has taken place.

He argues that factors other than tracking are likely to affect the importance of family background in early tracking countries. Because his results are contrary to those found by other researchers, particularly Hanushek and Wößmann (2006) and Ammermueller (2005) who find that tracking exacerbates educational inequalities using some of the datasets used in this paper, Waldinger investigates further the 'robustness' of his findings by carrying out a separate 'sensitivity analysis'. He attempts to replicate the analytical process followed in these studies and shows that slight changes in tracking measures (for instance, which age or grade is considered as indicating the occurrence of tracking) or in the sample used leads to different results and that, therefore, the results of these studies are not stable to slight changes in specification. 'These results', Waldinger maintains, 'cast serious doubt on the conclusions of a number of concurrent papers, such as Hanushek and Wößmann (2006), which find that tracking increases educational inequality and exacerbates the effect of family background for students' test scores'.

He states that further research needs to identify those factors that allow parents to improve a child's education leading to children from different backgrounds having unequal educational opportunities as 'even untracked systems give parents and students opportunities to select into better schools'. According to this study, caution needs to be taken in interpreting the relationship between tracking and the importance of parental background as causal.

3.4.7. *Jenkins, Micklewright and Schnepf (2006): Social segregation in secondary schools: how does England compare with other countries?*

In this paper, Jenkins, Micklewright and Schnepf attempt to identify levels of social segregation (defined as 'the uneven distribution across schools of children from different socio-economic backgrounds) in English secondary schools and draw comparisons between England and the situation in 27 OECD countries. They also compare levels of social segregation in England with those in Scotland and Northern Ireland as these two UK nations have different educational systems from England's. They argue that social segregation is an interesting area of study for a number of reasons. Firstly, concerns have been expressed about whether reforms to school admissions policies such as the 2005 White Paper on education which emphasised greater parental choice and greater independence of schools, led to increased unevenness in the social composition of schools. Secondly, if the peer effect is strong in determining other children's performance, academic achievement and later-life outcomes may suffer as a result of greater inequality.

The data used are obtained from the 2000 and 2003 rounds of the Programme of International Student Assessment (PISA) which collects information about 15 year old children and their schools. Measures of parental occupation used were parental occupation leading to the child's index of socio-economic position as being high or low. Another social background variable used was

based on the education of the child's mother. The authors' findings led them to the following conclusions:

England is a middle-ranking country – it comes near the middle of the distribution of social segregation in relation to the 27 OECD countries it is compared with. The highest segregation countries include Austria, Belgium, Germany and Hungary in contrast to the low-segregation Nordic countries and Scotland.

Little of the social segregation in English secondary schools can be attributed to the existence of private schooling. It was found that England's segregation was mainly driven by the uneven spread of children from different social backgrounds within the state sector. The authors, therefore, focused on the state sector in their analysis.

Parental choice in the state sector in England is high from a cross-national perspective – 52% of children in state schools claim that they attend their school because it is 'known to be a better school than others in the area'. However, differences in parental choice across countries do not seem to be strongly related to differences in social segregation levels.

The prevalence of school choice was found to be low: '28% of pupils in English secondary schools are in schools that use academic ability or feeder school recommendations as a criterion for admitting pupils, a level that is only half the average for all countries in the study (56%)'. The authors state that even this relatively low percentage may seem surprisingly high for a country whose school system is predominantly comprehensive and they emphasise that the School Admissions Code in the state sector (DfES, 2003) grants some discretion on comprehensive schools to select on ability. Higher levels of segregation were generally found in countries with a higher prevalence of school choice.

Several high-segregation countries have separate school tracks for academic and vocational schooling. In countries such as Austria, Germany and Hungary, over half of their high social segregation is accounted for by unevenness in the distribution of social background between the separate school tracks rather than within each of the school tracks. However, the authors claim that the PISA-based measure of school choice may not distinguish important country-specific aspects of school choice and the school system itself and, therefore, its power to account for the level of social segregation is weak. Similar limitations may also hold for the measure of parental choice. These findings, therefore, need to be interpreted with caution.

4. Limitations of attempts to evaluate selective systems

This chapter outlines some of the problems that arise in trying to evaluate the effects of selective systems. We make the assumption that one starts with no preconceived view about whether selection is good or bad, but simply wants to collect together the best available evidence on which to base a judgement.

No consensus has emerged from the empirical studies presented in Chapter 3. All of these studies have their limitations and the results from any one of them must thus be treated with caution. The reader could be forgiven for wondering if the question 'what are the effects of selective systems?' could ever be resolved by 'the facts'. It seems that one could almost take any arbitrary claim about the effects of selection and by searching hard enough find some study whose conclusion it was. Alternatively, if one were conducting one's own analysis of a given dataset, by making suitable assumptions, all of them quite reasonable, one might well arrive at a result in line with such a claim.

We argue that interpreting the evidence about the effects of selection will always be problematic and that the data do not speak for themselves on this matter. We list here a number of statistical and methodological reasons why the conclusions from existing research must be treated with caution. In Chapter 8 we will in fact conduct our own analysis of a national pupil dataset for England and indeed show that different choices of assumptions and models do lead to quite different conclusions.

4.1. Inability to control for other differences

In any kind of impact evaluation the researcher's task is to separate the effects of the programme, strategy or system being evaluated from the effects of other things. There are various evaluation designs that attempt to make this kind of separation possible, and various threats to the validity of inferences drawn from them (e.g. Campbell and Stanley, 1963).

In the case of research on selective systems, the challenge is to show not just whether there is a difference in academic performance, attitudes or social outcomes between children educated in selective systems and those in comprehensive systems, but that these differences arise directly as a result of the system, rather than from any other reason. Of course this is impossible to do directly and with complete confidence; the researcher's task has always been to approximate as closely as possible to a reliable answer, minimising the various threats to validity. However, we believe that the impact of these threats has been generally underestimated and the steps taken to overcome them insufficient.

Some of the justifications for this belief are quite technical and complex; they may also be controversial and would no doubt be disputed by others. However, we do believe it is important to the general argument about why existing studies are unreliable and therefore present a summary here. What the authors of the studies reported in Chapter 3 have largely done is to take a

complex social system, that is a school system with an input at age 11 and an output at age 16, and attempt to represent the reality of the system with a statistical model. In setting up such a model, assumptions have to be made which might be too broad, inadequate or just wrong.

In trying to separate the effects of selection from the effects of all the other aspects of secondary education and the effect they may have on academic attainment, it seems self evident that we must try to obtain a measure of these aspects and adjust for the effects they may contribute. We should then be left with the true effect that is attributable to selection. There are at least two reasons this may prove impossible in practice. The first is that in order to adjust for the effects of other factors, we need to know what they are and to have measures of them available. The second is that even if we include measures of all relevant variables, these measures will be imperfect and there will always be a level of error, or unreliability involved. In particular, lack of reliability leads to bias in the analysis. We discuss each in turn.

4.1.1. Key variables missing from the model

How do we know which variables to include in the model? There may be some obvious choices, without which a comparison would seem evidently spurious, though it may be surprising how often even these factors may be left out. Some of the earlier research on selection (e.g. Marks, 1983) simply looked at attainment outcomes of pupils in selective and non-selective systems, making no adjustment for any differences between these pupils that may have existed before they started school. Similarly Jesson, 2000, in his studies made no allowance for socio-economic differences in the pupils. Later analyses included measures such as pupils' Key Stage 2 levels as a covariate (i.e. a factor to be adjusted for) and it is clear that a good prior attainment measure will account for most of the variation in academic outcomes (Gray *et al*, 1986, Cuttance, 1994; Teddlie and Reynolds, 2000). However, if we were to make a list of factors that might affect academic attainment we would probably want to add others such as parents' socio-economic status, support for learning in the home, aspirations (of pupils and parents), academic self-concept, first language, ethnicity and gender (e.g., Nash and Harker, 1998; Teddlie and Reynolds, 2000). Very few research studies will have had access to measures of more than a few of these. Indeed, even prior attainment data, matched at the level of the individual pupil, became available for national datasets in England only very recently, (from about 1998).

In practice, the researcher's approach to this tends to be pragmatic. They will use any data available. Thus for example, Schagen and Shagen 2003, in using national data sets had to use what was available from the DfES, at the time. They said that "performance of a pupil (however defined) was assumed to be directly dependent on: prior attainment; percentage FSM in the school; sex; age; size of school; type of school; size of year group; percentage of selection in the LEA." They noted that there is evidence that social deprivation is strongly related to prior performance and school type, and so it is essential to include this in the model, but the percentage of FSM in a school was the only measure of deprivation available to them.

4.1.2. Bias arising from reliability of covariates

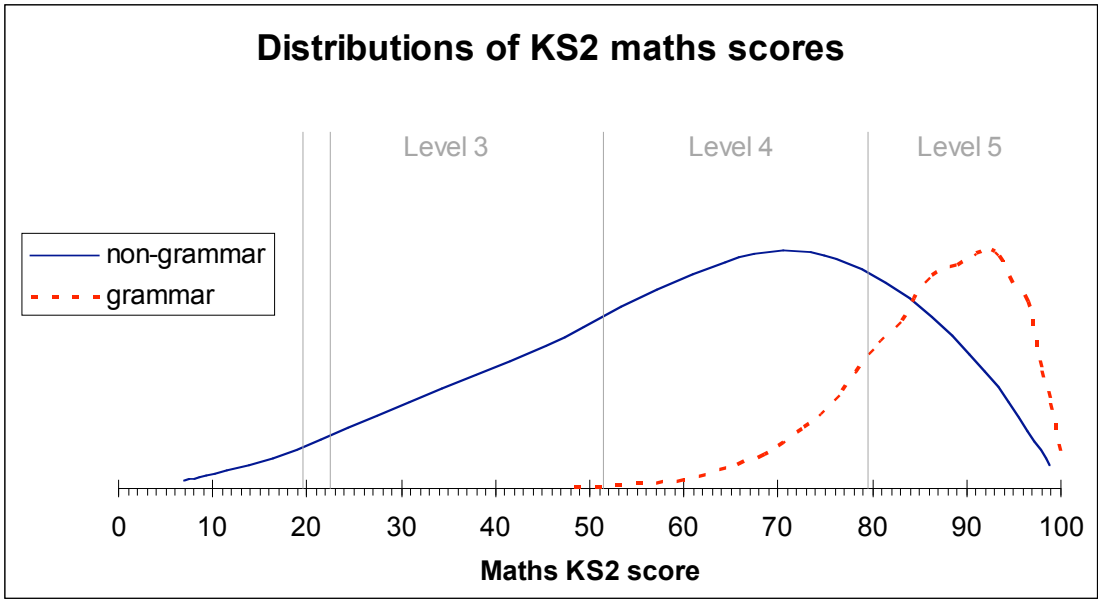
The second reason is that even if we know what factors to control for, and have measures of them available, those measures will be imperfect. To take

an example, suppose we have a measure of prior attainment, but it is only a broad level rather than a raw mark. In the case of KS2 data, this is in fact commonly the case since almost all pupils attain levels 3, 4 or 5 and marks are often not available. Again, for example, Schagen and Schagen, 2003, noted that marks were not available for Key Stage data, but justified the use of levels arguing that marks would only be of interest for individual pupils, whereas they were looking for whole school effects. Let us suppose also for the sake of argument that if we could adjust our outcome measure (say a points score at GCSE) for the effects of the raw mark at KS2, the variation that remained would be entirely due to the school and could therefore be interpreted directly as an indicator of effectiveness. In this case the effect of using broad levels instead of raw marks is that we can explain less of the variation in outcomes; the grouping of a spread of marks into the same level loses information and makes our baseline a less good predictor of future attainment. This is because there is now an unknown error in the baseline for each pupil, (level + error = raw mark), which in turn introduces an error into the GCSE points score, and so into the measure of school effectiveness. Thus using levels rather than raw marks at KS2 we still give us a measure of school effectiveness, but with errors, or bias, introduced. Thus if there is a school whose pupils' raw KS2 scores are generally towards the top end of a range for which a level is defined, this school will appear spuriously more effective; the pupils started from a higher baseline than indicated by the KS2 level awarded to them and a bias has been introduced.

This situation is illustrated in Figure 2, which shows the distribution of marks achieved at Key Stage 2 in maths by the 1999 cohort, for both grammar school and non-grammar school populations.[3] It also shows the usual cut off-points for levels to be awarded. Not surprisingly, the majority of grammar school pupils achieved level 5. What is striking, though, is that if one considers all pupils who achieved, for example, level 5, those in grammar schools typically have significantly higher marks than those in other schools. In fact the averages are 89.4 and 86.8 respectively. It would not be surprising, therefore, if level 5 pupils in grammar schools went on to achieve higher GCSE grades than level 5 pupils in other schools. An adjustment that controls only for the level achieved at KS2 will have failed to control for a significant difference between them and will make the grammar schools look better than they really are.

[3] By 'grammar school' and 'non-grammar school' populations we mean those pupils who subsequently went to each type of school. The 2001 cohort are those who went on to take GCSEs in 2006 and form the subject of our main analysis. Note that the way that QCA awards levels is imprecise. At any level achieved there are pupils with marks outside the ranges shown, though the vast majority are inside. Level 6 is awarded on the basis of an additional paper, so is not directly related to marks achieved on the main test.

Figure 2: Distribution of KS2 maths scores, with associated levels, for grammar and non-grammar school pupils



It might seem obvious that the solution to this is to use marks rather than levels, and this is indeed to be preferred. In Chapter 8 we do just this and show that doing so does indeed reduce the apparent benefits of attending a grammar school. None of the studies we have reviewed, however, appears to have made use of marks, probably because they were not available.

It might be argued that KS2 scores, whether as raw marks or levels, are not good predictors of GCSE outcomes five years later. It may be that marks on an entrance test at age 11, based on cognitive abilities, would be a better predictor. However, for this to have validity the same test would have to be taken by all children. Interestingly such a measure was included in the NCDS data.

Unfortunately, though, even using marks does not completely solve the problem. The same bias arises whenever any baseline measure has less than perfect reliability or precision. Sadly that means it arises all the time, since no measure is ever perfect.

This difficulty of being unable to adjust properly for initial differences between pupils is a case of the problem of 'regression to the mean'. This is due to the random error in both the KS2 score and GCSE score. Fortunately, it is possible to correct for the reliability of the baseline measure, here KS2 score, and we present such corrected analyses in Chapter 8. Again, so far as we can tell, no previous study has done this.

4.1.3. Important differences between selective and non-selective systems

From the above it is clear that making confident comparisons of the effects of different systems is quite problematic, especially when the contexts in which the systems operate are themselves different. It is therefore important to know what the differences are between the two kinds of system. There are three levels at which differences between selective and non-selective systems

can appear: at the level of the area affected by selection, at the level of the school and at the level of individual pupils.

The first of these is itself problematic, since how does one define the area affected by selection? Most studies have used the LEA as the unit of analysis (e.g. Jesson, 2000; Atkinson *et al*, 2004; 2006), though as we discuss below, this is a rather crude approximation. Indeed these authors had to make decisions about what level of percentage of pupils attending a grammar school, would constitute a selective LEA. The majority of grammar schools are in LEAs that are not wholly selective. Many of them recruit students from a wide area. Even where LEAs are wholly 'selective' or 'comprehensive' in their provision, the number of students crossing LEA boundaries to go to school may be higher than has previously been acknowledged.

In our analysis (Chapters 7 and 8) we have tried to identify geographical areas that are affected by selection. It has not proved possible to draw a border between two areas such that we can say one area is affected by selection and the other is not. However, it is clear that in general there are some differences between areas that are subject to selection and areas that are not, particularly the general level of affluence in that area. For now we just note that the existence of such differences makes it difficult to know whether any apparent differences in performance are due to the existence of selection or due to other factors (see Section 4.5 for more detail).

The schools themselves may also be different. For example, grammar schools may be more likely to be single sex, smaller, and to have a sixth form than other schools. Selective areas have also been shown to have a higher proportion of faith schools (Atkinson *et al*, 2004). Again, we present more detail on the differences in Chapter 6.

The problem here is that if any of these factors is itself associated with pupils making greater progress, it will be hard to untangle the effects of that factor from the effects of being a grammar school. For example, if on average single sex schools are more effective than mixed sex schools it could be that the apparently greater effectiveness of grammar schools when they are compared with others is due to their single sex status. Although it might be true on average that grammar schools are more effective than other schools, this would be more a result of their tendency to select by sex than of their selection by ability. On the other hand, if we control within the model for the effects of being a single sex school, we might be inadvertently removing some of the genuine effects of being a grammar school, since it could equally well be the other way round: the apparent advantage of single sex schools comes from the high proportion of grammar schools among them.

Statistically, one can solve this conundrum by looking for interaction effects. Is there an effect of being a single sex grammar school, over and above that of the separate effects of being a single sex school and of being a grammar school? Unfortunately, none of the studies we have reviewed has reported looking for these kinds of effects. Shagen and Shagen 2003, did include some interaction effects in their multi-level model; these were between prior attainment and other explanatory variables. They were not reported on explicitly and they did not consider single sex schools.

The third level of difference between grammar schools and others is the pupils themselves. One might speculate that those who go to grammar schools, as well as being more able, tend to be more motivated, have more

support for learning in the home and have higher aspirations. If so, we should expect them to achieve more at KS4 and beyond, whether or not the grammar schools are 'effective'. Unfortunately, all these differences are only speculative, since no data are available within the national pupil datasets, apart from KS2 scores and eligibility for free school meals.

4.2. *Problems with the quality of baseline data*

There are a number of problems with the availability or quality of baseline data.

4.2.1. KS3 is not a baseline, nor is it an outcome

Some comparisons of the effectiveness of different types of schools have made use of datasets in which Key Stage 3 results were used either as the baseline for KS4 or as an outcome measure with KS2 as a baseline (e.g. Jesson, Schagen and Schagen). These are clearly somewhat unsatisfactory, as the authors themselves often acknowledged. The results of these studies therefore need to be treated with caution, as noted in Chapter 3.

4.2.2. Key Stage 2, especially just levels, is not an adequate baseline

We have already mentioned the problems of using broad levels to equate groups that are as different as grammar and non-grammar school populations (see 4.1.2, p108). Other objections to the use of KS2 tests as a baseline include that they measure achievement in only the three subjects, namely mathematics, English and science. (Prais, 2001 made the same observation for KS3). If KS2 is only available as levels rather than marks, one could also criticize its use on the grounds that its range is too limited. Level 6 is not available to most pupils.

4.2.3. Key Stage 2 may be affected by selection

It is certainly possible that KS2 might be treated differently in primary schools in selective and non-selective areas. For example, in an area where entry to a grammar school is seen by many as the most important outcome of primary schooling, it could be that the emphasis given to KS2 would be correspondingly lower than in areas without any selection. If this were the case, KS2 results in selective areas would be generally lower, other things being equal, and hence value-added calculations from KS2 to KS4 would make pupils in these areas appear to make more progress.

On the other hand, it is possible that pupils who are offered a place at a grammar school receive such a boost to their self-confidence and motivation as a result that they go on to excel themselves at Key Stage 2 and achieve higher scores than they would otherwise. In this case, value-added calculations might be said to make grammar schools appear to be less good than they really are.

4.2.4. Free school meals data are problematic

Eligibility for free school meals (FSM) is the measure that it most widely used to indicate socio-economic status. Although some have argued that this is a

good measure (e.g Gorard *et al*, 2003, p40), there are a number of problems with it.

First is a concern that differences in the rates at which children with different cultural backgrounds are likely to identify themselves and claim their free school meals[4] makes it problematic to use as an explanatory variable in judging the relative performance of different types of schools. Dietary requirements, the desire to avoid stigma or not to receive charity could all be different for different cultural groups.

The fact that FSM is a dichotomous variable at the individual pupil level also limits its explanatory power. Pupils are either eligible or they are not, whereas the underlying construct of social advantage in relation to educational opportunity is likely to be more of a continuum, with degrees of disadvantage needing to be differentiated. The fact that around 16% nationally are eligible for FSM means that this variable cannot distinguish between, for example, a student who is average and one who is highly advantaged socially; neither is eligible for FSM.

We can illustrate the deficiency by comparing the proportion of variance in GCSE maths grade that is explained by the free school meals yes/no variable (4.6%) with what can be explained by a slightly more sophisticated measure of socio-economic status (SES) from the Yellis project.[5] This latter variable is a scale measure based on students' reports of their parents' occupations and levels of education and accounts for 9.8% of the variance in GCSE maths, over twice as much as the simple FSM dichotomy. Part of this difference is due to the fact that a dichotomous variable will not account for as much variation as a continuous one, but even when the Yellis-SES measure is dichotomized at the median value, it still accounts for 6.7% of the variance.

We were not able to match FSM and Yellis-SES data at the individual pupil level, but could do this at the school level. When the outcome is a school's percentage achieving 5+A*-Cs, % FSM accounts for 58% of the variance in a quadratic regression model, compared with 69% of the variance for average Yellis-SES. Once again, the Yellis variable seems to be a better measure of educational disadvantage than free school meals.

Of the existing studies, only those based on the national birth cohorts (Steedman, Sullivan and Heath, Galindo-Ruenda and Vignoles, Manning and Pischke) have been able to use high quality SES measures which are likely to be comparable to the Yellis measure. Those studies that have used national pupil datasets, (Schagen and Schagen; Atkinson *et al*, Levačić and Marsh) have had only free school meals as a measure and are therefore subject to the limitations outlined above.

4.3. *Problems with the quality of outcome data*

4.3.1. *Ceiling effect at GCSE*

A real problem for measuring the achievements of the most able pupils is the limitations of the top end of the GCSE scale. For the population as a whole,

[4] Although the expression used by the DfES is 'eligible for free school meals', a pupil or their family generally has to identify themselves to the school as having this eligibility, though not necessarily eat the meals, in order to be counted as 'eligible'.

[5] See www.yellisproject.org

the top grade, A*, is relatively rare, but for pupils in grammar schools it is far more common. For example, in maths GCSE in 2004, 2.3% of pupils in non-grammar schools achieved A*, compared with 22% of those in grammar schools. It seems reasonable to assume that if an even higher grade existed (say an A**), a fair proportion of grammar school pupils would achieve it, but only a tiny fraction of those in other schools would. If this A** grade did exist then any analysis that simply ignored it and treated all A**s as if they were just A*s could significantly underestimate the achievements of grammar school pupils and certainly shrink the apparent difference between them and pupils in non-grammar schools.

Unfortunately, as we have GCSE data only as grades, the analyses that exist or that we can do with existing data are exactly equivalent to ignoring the higher A** grade, since it does not exist.

There is a further aspect of this problem of the failure of GCSE to discriminate at the top end of achievement. Schools with significant proportions of the most able pupils might decide that taking GCSE examinations would not stretch them adequately and so not enter pupils for these exams. Alternatively, they might enter pupils early, perhaps a year or more ahead of their peers, at a stage when the examination would be appropriate to the stage they had reached in their school studies. These early entries might do well, but in many cases would not do as well as they would have done had they waited. In either case, any analysis that compared GCSE performance of different types of school would underestimate the achievements of pupils in grammar schools. However, early entry and missing GCSE and moving straight on to A-level, are likely to be practices of the past, as pressure to be seen to do well in the performance league tables has forced schools into entering their pupils conventionally.

4.3.2. Interval scales

Some studies (e.g. Schagen and Schagen, 2005) have pointed out the arbitrariness of coding GCSE results using particular scales. For example, it is common to convert grades to points, using A*=8, A=7, B=6, … G=1, U=0. However, there is no reason in principle why one should not allocate these points differently, and this could make a significant difference to the results. If we were to award A*=9 and leave the other grades the same, for example, this would make far more difference to the grammar schools than to other schools and hence make the former seem much better.

The issue here is not the exact numbers used to represent each grade, but the relative intervals between them. The scale used by the DfES until 2004 (A*=8, A=7, … G=1, U=0) would give identical results to the new QCA points score (A*=58, A=52, B=46, …G=16, U=0, i.e. 6 points between grades) if it were not for the relative gap between G and U changing from one grade (one point) to being nearly three grades (16 points). This effect is not likely to be very great, but it would require further investigation into the awarding of the U grade and level of entry at GCSE to establish just what the effect is.

4.3.3. Use of 'percentage reaching threshold' outcomes

Many studies have used the percentage of pupils in a school, or type of school, who achieve a particular standard as a way of quantifying the outcome. For example Jesson 2000, used the usually accepted standard

measure of 5 GCSEs awarded at grades A* to C. Though this may have the advantage of being simple to understand, and avoids the problem of assuming interval scales, it has a number of statistical disadvantages, particularly when it is used to compare the achievements of groups that are initially quite different (Schagen and Schagen, 2005). There is also a certain arbitrariness about it; there is no statistical reason to chose 5 GCSEs, it has simply become an accepted benchmark. Also for GCSE, the biggest criticism of 5 A* to C grades is that there is no discrimination between subjects; all are assumed to be of equal difficulty and also to have equal accessibility to all pupils.

4.3.4. Subject difficulty

The question of whether some subjects are 'more difficult' than others is a controversial one (Fitz-Gibbon and Vincent, 1994, 1997, Goldstein and Cresswell, 1996; Sparkes, 2000; Coe *et al*, 2008). What is beyond doubt is that the profiles of students entering different subjects are different and that the average grades achieved by the same students in different subjects also varies. If the subject difficulties argument is accepted, it follows that students in different types of schools are likely to take different combinations of subjects with different difficulties, and that there may be systematic differences in the value that should be attached to what are apparently the same grades achieved in different subjects.

If, for example, pupils in grammar schools typically take 'harder' – or more severely graded – subjects, analysing their performance using total points achieved, average points, average of the best eight, or the percentage gaining 5+ A*-Cs, would underestimate the true difference between their achievements and the achievements of pupils in other schools who typically take 'easier' subjects. Although all pupils are required to study the core subjects of the National Curriculum they do not necessarily have to sit an examination; there is scope for wide variety in the actual subjects taken in schools and more so in comprehensive schools where the ability range is wider. Also other qualifications, such as short-course GCSE and GNVQs have been allocated point scores by the DfES, with an unjustified equivalence. None of the studies we have reviewed has acknowledged this issue, which we return to in our own analysis in Chapters 6 and 8.

4.3.5. What is the outcome of secondary schooling: GCSE, A level or beyond?

Most of the attempts to compare different types of school in England have focused on Key Stage 4 as the outcome. There are some obvious reasons for this, such as the availability of data and the argument that the years of compulsory secondary schooling (age 11-16) represent a natural and self-contained stage. Also, it would be quite problematic to use any later stage, since a fair amount of mixing takes place at 16+ to the selection that may have occurred at 11+: many pupils change school or go to college at 16+ and sixth forms are themselves often quite selective.

Nevertheless, in a grammar school where the majority of pupils go on to take A levels and then go to university, GCSE might well be seen as no more than a stepping-stone to these more important outcomes. Even in schools where the routes pupils follow are more mixed, it could be argued that in terms of pupils' life chances and quality of education, outcomes other than GCSE are the most important.

Such outcomes do form a part of the comparisons based on the birth cohort studies (Steadman, Galindo-Rueda and Vignoles). However, the number of pupils involved was so small that these did not have any significance. However, Yang and Woodhouse, 2001, focused on AS/A2 attainment in their study, and found that on attainment alone, pupils at grammar schools were about 6 A-level points better than those at comprehensive schools.

4.3.6. Lack of evidence about non-academic outcomes

Related to the previous point is the lack of evidence about the effects of selection on outcomes such as enjoyment of schooling, self-esteem, socialization and social skills, among many other attitudes and behaviours that might be valued. Even within the area of academic outcomes, one could argue that outcomes such as the capacity for independent learning, intellectual curiosity and critical thinking are more important than GCSE grades. Needless to say, none of the studies evaluating selection has anything to say about any of these things.

4.4. Issues in the calculation of value-added

4.4.1. Use of individual pupil-level data

Some of the earlier studies reviewed in Chapter 3 could not make use of individual pupil-level data, since such data were not available. These included the work by Marks *et al* (1983, 1985). There has been considerable debate around the question of the adequacy of using such data for evaluating school performance (Teddlie, Reynolds and Sammons, 2000, p74; Gray, Goldstein and Jesson, 1996) Not least amongst the problems is that across a cohort of pupils in a year group there is likely to be considerable variation in any measure. It is notable that in the models reviewed in Chapter 3, the variation in outcome at GCSE is mostly accounted for in variation at the pupil level. However, all this variation gets aggregated up into gross statistics that represent the school, either as a Key Stage 2 average, or GCSE outcome average.

4.4.2 Different models

Discussion in Chapter 3 and also in this Chapter has highlighted that researchers need to make decisions about how to set up their statistical models, and in particular what explanatory variables to bring into a model. This raises the question of whether the more recent development of multi-level modelling is a superior technique to the well established OLS multiple regression technique. Similarly should whole school variables be included in models as well as pupil level models. We had to make decisions like these when developing our own models; the development of these models and discussion of the issues involved in making the decisions are left to Chapter 8.

4.5. Wrong unit of analysis

If we are to try to estimate the impact of selection, we need to identify which groups to compare. At first sight this may seem a simple enough issue: compare those areas in which there is selection (grammar schools and secondary moderns) with those in which there is no selection (comprehensive schools). For many of the studies we have reviewed, the Local Authority (LA)

provides an apparently simple answer to the problem: some LEAs are selective, some are not.

Unfortunately, there are also some that are partly selective, with a mixture of grammar, secondary modern and comprehensive schools in the same LA. Indeed only three LAs contain no 'comprehensive' schools. In the partly selective LAs the number of comprehensive schools varies between 1 (Bournemouth, Slough) to large numbers in big shire counties (Essex 77, Lancashire 82). Further detail is given in Section 1.1.3 (p10). It seems clear that the designation of a school as 'comprehensive' or 'secondary modern' is somewhat arbitrary and may not reflect the true selective status of that school. This problem casts considerable doubt on the many analyses that have used these school categories. We noted in Chapter 3 how some authors (Jesson; Atkinson) assumed that once they had defined their selective LEAs, all schools within them were either grammar schools or non-grammar schools, treating secondary modern schools and comprehensive schools as the same type of school.

Related to the problem of categorizing schools is a question about the extent to which an LA can be regarded as an intact unit. Imagine, for example, a selective LA neighbouring a non-selective one. Pupils who live close to the border may well not stay within their own LA. Those in the non-selective area may take the 11+ and, if they pass, cross the border to attend a grammar school. Equally, those in the selective LA who fail the 11+ might choose to cross the border to a 'comprehensive' school rather than attend the local 'secondary modern'. It should also be noted that LAs vary greatly in geographical size. The recently formed unitary authorities are based around cities and are thus small geographically when compared to large shire counties. This will clearly affect the amount of border crossing that can realistically take place. There may, of course, be other reasons why pupils might go to a school out of their LA. It is certainly possible in this scenario that it will be the relatively socially advantaged pupils in both cases who are most likely to choose and be able to get to the further away school. A comparison of selective and non-selective LAs here would tend to exaggerate the benefits of grammar school attendance and diminish the advantages of the secondary moderns, purely as a result of ignoring this cross-border traffic.

In fact, our analysis in Chapter 6 shows that nationally around 8% of secondary age pupils cross an LEA border to go to school, while for those attending grammar schools the figure is 20%. These numbers are certainly large enough to undermine the validity of using the LA as a unit of analysis of the effects of selection. None of the studies we have examined seems even to have acknowledged the possibility of a problem here, let alone attempted to address it.

4.6. *Heterogeneity of selective systems*

It is clear from the above that even defining an area as 'selective' is problematic. However, it is also apparent that many different kinds of selection can operate.

Different selective systems are very different and may have very different effects. Most studies have not even tried to investigate any differential effects, e.g. with respect to the level of selection operating. The tendency has been to

recognise that there is variation and then treat the selective areas as if they are homogeneous. We illustrate in our own analysis in Chapters 6 and 8 the difficulty inherent in defining level of selectivity.

4.6.1. Differences in the selection process

The number of places awarded to children in grammar schools clearly depends directly on the number of places available. The amount of choice parents have in applying to a grammar school if they decide to, clearly varies across the country. The amount of choice depends on the number of grammar schools in particular areas and to what extent the parents and/or pupil are willing to travel. How this actually affects decisions to apply, or not, is again worthy of further study.

In virtually all LAs where selection operates it is an opt-in decision by children and their parents; the only exception we are aware of is Buckinghamshire. Typically parents who wish their child to attend a grammar school have to go through a complex process of both applying to the school and applying through the school preference procedures of the LA. Although there is guidance available on how to do this both from schools and LA officers, parents have to make a conscious decision to apply. Others may have views that are philosophically opposed to selection or may find it too much hassle to attend open evenings and deal with all the paper work, so an application is not made on behalf of their child no matter what his/her ability. This is all speculative. We have no information on school entrance tests, who takes them and is awarded a place or is not awarded a place, and who might take them but chooses not to, and who is totally unaware of the opportunity. This is clearly an area worthy of further investigation, but it cannot be pursued further in this report.

4.6.2. Differences in the level of selection

Another key difference is in the position at which the selection cut-off is located. We have highlighted the wide variation in the level of selection in 'selective' LAs across the country (see Table 1, p12). We have found some evidence in the performance tables that a schools achievement is related to the level of selection in the LA: the 2004 DfES Performance Tables show a general trend that the higher the level of selection, the lower the school is placed, illustrated in Table 26 (percentage selection from Table 1).

Table 26: Comparison of percentage selection in the top performing schools in England

	Top 50 performing schools nationally	Top 100 performing schools nationally
Percentage selection for the LA	Number of grammar schools	Number of grammar schools
1-10%	25	32
11-20%	10	17
21-30%	12	30
Over 30%	2	14

We can broadly see that where selection is at its lowest (i.e. grammar schools are only creaming the top 10% or less) schools are more successful in attaining a high position in the performance tables. Although the top 100

performing state schools are dominated by grammar schools, those in LAs with lower selection levels tend to be placed higher.

This raises several difficult questions concerning whether there is an optimum level of selection. We return to this in the analysis of our own data in Chapters 6, 7 and 8.

4.6.3. Key differences among 'comprehensive' systems

In areas where there are no grammar schools we have already discussed in Chapters 1 and 2 (1.1.2, p8 and 2.1.7, p18) that other forms of selection take place. This may be by postcode in that the price of housing in the proximity of a *good school* may rise to a point where the school becomes inaccessible to many parents who otherwise would have it as first preference for their son or daughter. Faith schools usually require some evidence that those applying and their parents are active practitioners in the faith concerned. It has been suggested that some families start attending church as their son or daughter approaches the end of their primary schooling. There are also many single sex schools; including 122 of the 164 grammar schools. Specialist schools can select up to 10% of their intake by aptitude. Whether aptitude is somehow different to ability is an interesting questing, which we will not pursue further here (West, Hind and Pennell, 2004).

Independent schools are also a prominent feature in the selection debate. Many of these schools select by ability themselves, and all of them are selective in that the fees need to be paid, which usually means affluent parents. An interesting question to pursue is how many children who apply to a maintained selective school but fail to secure a place then go on to an independent school. Similarly in some areas of England there are state maintained independent schools, the City Technology Colleges and Academies. We know of no research into how the presence of independent schools in an area affects selection.

4.7. Focus on cohorts educated in the 1970s

The process of comprehensive reorganisation that took place in England during the 1960s and 1970s must have caused enormous disruption to many schools. Schools were often amalgamated, though some were closed. Some staff will have lost jobs or had significantly different roles imposed on them. All of this might have been unsettling to the education of those students going through the system at the time, and could well have affected their progress and attitudes. On the other hand, it could be that the opportunity for a fresh start allowed stale practices to be re-evaluated and schools to be invigorated. Idealistic new teachers may have been inspired by the egalitarian ideals of comprehensive education and brought fresh enthusiasm to the new schools, thus giving a boost to the experience of students in comprehensive schools that would not be sustained.

For the studies conducted during the 1970s and 1980s, it is possible that any such changes could still be working through the system. Comparisons of selective and non-selective systems may therefore tell us more about the management of this enormous change than about the stable and generalizable longer term impacts of different ways of organising schooling.

Although these changes are now largely a distant memory, and the number of grammar schools has remained stable since 1999, we need to remember that several contemporary studies have used data from the same 1958 birth cohort study. People born in 1958 will have typically entered secondary school in 1969, just as the reorganisation was coming to its peak. It could certainly be argued that this cohort was something of a special case and their experiences may not be typical of students who passed through the system later. We have made such criticisms in Chapter 3. It would inform the debate if there were a more recent birth cohort longitudinal study, but we found no such study reported in the literature.

Although the number of grammar schools has been stable for 8 years, this is against a background of continuous change in legislation concerning the structure of secondary education (Chapter 1). Thus in 2008 we cannot really be sure what type of schools we are comparing the maintained selective schools with; at best we can call them non-grammar schools.

4.8. Researcher bias, assumptions and selective presentation

It might be nice to think that detached, objective researchers would collect data about the effects of selection, analyse them neutrally, and dispassionately report their findings. Even the most casual acquaintance with the research on selection would show that this is not how it is. The issues are so politically charged and commitment to existing beliefs appears so strong that the data can seem almost irrelevant. In this debate, evidence is more often used as a garnish or rhetorical flourish than as a foundation for decision making. Is it possible that the elaborate and sophisticated analyses reported in the literature are no more than a front for pure prejudice?

An interesting example of the tendency of research to confirm rather than disprove prejudices comes from the field of expectation effects. Studies have shown (e.g. Rosenthal, 1966) that even for a relatively simple and 'objective' measure such as the speed a rat learns to negotiate a maze, results can be affected by inducing particular expectations on the part of the researcher. In effect, if the researcher believes that one rat is cleverer than another, the scientific evidence will tend to confirm this, even when there is no genuine difference. Rosenthal discussed what he calls interpersonal expectancy effects and teaching in the classroom; if a teacher expects a child to do well then there is a tendency for the child to do so. Rosenthal's research has found these interpersonal effects in a wide range of human activity. It could well be the case that if researchers believe a particular school structural system to be better than alternatives, then their research findings will support this view. We note from the literature review that both Marks (pro grammar school) and Jesson (pro comprehensive school) appear to have overstated the interpretation of their results.

In any empirical research there is an element of selection. Some data are collected, others are not; some methods of analysis are chosen, others rejected; some results are reported, others ignored. All of these decisions rest on value judgements which are to some extent arbitrary. It would not be possible for a researcher to have taken such decisions without some sense of how their choices might affect the final results, however much they may genuinely seek to be balanced and neutral. If the researcher were not even

seeking to be balanced and neutral it would often not be difficult to use such selection to present a particular story.

In Chapter 8 we list a number of different choices a researcher must make in analysing the same dataset. This list is by no means exhaustive and was to a large extent limited by the time available to conduct the analyses. Nevertheless, even these limited choices give rise to over 500 different answers to the question about the effectiveness of grammar schools. For now we simply note that few of the studies we have reviewed offer more than one answer to the question. In determining the effects of selection in education, selection in presentation of results is also important.

5. Discussion of the evidence from the literature review

Anyone who has read Chapter 3 is likely to be left with a sense of confusion over what the evidence shows. Different studies, conducted at different times, by different researchers using different methods appear to have come to quite different conclusions about the impact of selective education. Reading Chapter 3 will not have helped this feeling of bewilderment. It will be apparent that many of the methods used are claimed to be biased or inappropriate and their results suspect. What then, if anything, can we conclude from the review?

In this chapter we attempt to summarize the key messages of Chapters 3 and 4, and to draw what conclusions we can.

5.1. Summary of the literature review

We present here the key findings of the main studies reviewed in Chapter 3.

Steedman (1980) used data from the National Child Development Study (NCDS) of individuals born in a particular week in March 1958. These pupils attended secondary schools during a period of significant comprehensive reorganization. No clear differences were found between academic attainment in selective and comprehensive systems, after allowance for initial differences. Behaviour was perceived by teachers to be worse in comprehensive systems. In a follow up study, Steedman (1983) found that the highest attainers in selective systems were doing slightly better at O level and A level than those in comprehensive systems, though the result was reversed for lower attainers. Overall, given the limitations of her data, she concluded that there was no clear evidence that either system was better. A reanalysis of the same data by Kerchoff *et al* (1996) came to much the same conclusions.

Studies by Marks *et al* (1983; 1985) used aggregated O-level, CSE and A-level results of two samples of schools in England in 1981 and 1982. They found a general tendency for the most selective LAs to get the best results. Various adjustments were made to try to overcome limitations of the data, with no change to this finding. Nevertheless, significant limitations remained and the trustworthiness of the overall findings must be open to question.

A written answer in the House of Lords (2000) appeared to address the question of the comparison of performance of grammar and comprehensive schools using national data. However, we judge the methods and reasoning of this approach to be wholly unreliable.

A number of studies by Jesson (2000; 2001) used national data sets for 1998 GCSE performance. He found that comprehensive systems were outperforming selective systems. However, specific concerns and uncertainty over his approach make this finding questionable, and certainly the force with which it is presented seems unjustified.

Prais (2001) reanalysed the same data used by Jesson and criticized Jesson's approach, outlining a number of difficulties in conducting comparisons of this

kind. Although Prais' analysis put grammar schools ahead of comprehensives in the value-added for top ability pupils, he was reluctant to read any significance into this.

Jesson (2007) later investigated the issue of whether grammar schools offer children from poor backgrounds a 'ladder of opportunity' to reach higher levels of achievement than would be expected in comparison to being educated in state schools. Using free schools meals as an index of disadvantage and the income deprivation affecting children index, he shows, firstly, that there are few disadvantaged pupils educated in grammar schools in comparison to the national average in selective schools and, secondly, that there is a clear polarisation in grammar school intake towards choosing pupils from low disadvantaged communities. He argues that there is a cycle of deprivation and of limited opportunities that is perpetuated as grammar schools are 'ghettos of the advantaged' rather than providing a ladder of opportunity to children from disadvantaged backgrounds.

Yang and Woodhouse (2001) considered A-level results in different types of institution taken between 1994 and 1997. When a range of individual student level characteristics were controlled for, students in maintained selective schools performed better than students in comprehensive schools. However, when the academic composition of an establishment as measured by average performance at GCSE was included in the model, there seemed to be no difference between progress (value-added) at A-level in grammar and comprehensive schools.

Schagen and Schagen (2003) used national pupil data matched from KS2 to KS3 and another cohort matched from KS3 to KS4 in 2000. They used multilevel models to compare the performance of schools according to their type and LEA's level of selection. Their results showed some advantage to pupils in grammar schools, especially in KS3 performance for those at the borderline of grammar school entry, and at GCSE for pupils who were achieving at about levels 5, 6 and 7 at KS3. Other comparisons showed no difference or small advantage to comprehensive LEAs. This analysis seems to be a substantial methodological improvement on previous work, though the authors acknowledge some significant limitations of the available data.

With the availability of national pupil level data matched from KS2 to KS4 in 2001, Schagen and Schagen (2005) followed up their earlier work. Analysis of this cohort largely confirmed their previous results. A large grammar school effect was found for borderline pupils – those who narrowly gain a grammar school place – who obtained much better GCSE results five years later than pupils of equal prior attainment in comprehensive schools. However pupils with KS2 average level higher than 5 appeared to fare better in comprehensive schools. The number of such pupils is however, very small; about 0.25% of a year group.

Atkinson *et al* (2004) used the national pupil dataset and PLASC (pupil level annual school census) data for the cohort taking GCSE in 2002 to compare the performance of pupils in selective and comprehensive LEAs. Their analysis used OLS regression and propensity matching, a technique which aims to find matches, so that like is compared with like. The two kinds of LEA were matched on a number of variables to attempt to compare like with like. There was no overall difference between selective LEAs (defined as those with 10% or more of their pupils in grammar schools) and non-selective LEAs. Pupils at grammar schools in the former did better than comparable pupils in the non-

selective areas, but pupils who did not attend grammar schools in selective LEAs did slightly less well than comparable pupils in non-selective LEAs. This study also reported that poorer pupils, as defined by FSM (free school meals) status in selective LEAs were only half as likely to attend a grammar school as those with similar KS2 scores. However, those pupils from poor backgrounds who did attend grammar schools from age 11 did better at GCSE than might otherwise have been expected.

Sullivan and Heath (2002) returned to the NCDS 1958 birth cohort in a comparison of independent and state school performance. Within the latter, a division was made between grammar schools, secondary moderns and comprehensives. After controlling for intake to the schools, the authors found that pupils at state grammar schools and private schools achieved superior educational outcomes at age 16 to pupils at comprehensive schools.

Galindo-Rueda and Vignoles (2004) also used the NCDS data to estimate the impact of selection. They found that the most able pupils in selective systems did better on a mathematics test at age 16 than those of similar ability in comprehensive systems, and particularly so for girls. They found no significant difference for middle and low ability pupils.

Shuttleworth and Daly (2000) examined data from Northern Ireland and found that pupils in grammar schools performed substantially better at GCSE than comparable pupils in secondary [modern] schools.

Gallagher and McKeown (2000) examined the attitudes of pupils in selective schools in Northern Ireland and comprehensive schools in Scotland. Modest differences were found on some variables, no differences on others. In general, pupils in all types of school appeared to be positively disposed towards their own school.

Croxford (2000) investigated whether the differences between the four UK education systems were associated with different levels of social inequality in attainment, using data from various cohort studies. The study found that there was more social segregation between schools in England and Northern Ireland than in Wales and Scotland, and correspondingly more differences in attainment at school level in the former two nations. She found that there was no evidence that the overall levels of attainment were affected by the extent of selective or comprehensive schooling, but that standards across schools were more uniform in the two systems which have least differentiation between schools (Scotland and Wales).

Croxford and Paterson (2006) explored trends in social class segregation between British secondary schools for the period between 1984 and 1999 aiming particularly to identify any differences in trends in the more comprehensive school systems in Scotland and Wales compared to those in England. Using data from two cohort studies, some differences in levels of segregation between the national systems were found: segregation was lower for working class pupils in Scotland than in England and also in Wales for the 1980s cohorts and particularly for managerial and professional class pupils up to 1993. The authors argued that these two findings lend some support to the hypothesis that segregation is lower in more comprehensive education systems. No consistent increase in the level of segregation in the 1980s and 1990s was found. However, evidence of a 'starting-gun' effect and subsequent decline after the 1988 Education Reform Act was found lending some support to the hypothesis that parental choice legislation has increased

segregation. The authors though mentioned the limitations and the tentative nature of their data.

Marsh (1991) examined pupil attitudes in Australian schools. Pupils in high ability schools had generally more negative attitudes. Differences between educational systems in Australia and England are such that the relevance of these findings is questionable.

Hanushek & Wößmann (2006) used international data sets in reading, mathematics and science to identify tracking effects by comparing students' performance between primary and secondary schools across tracked and non-tracked systems. They find that early tracking increases inequality in achievement whereas the evidence about possible efficiency gains from tracking is mixed. Their findings have been challenged by Waldinger (2007) who argues that these are not robust to slight changes in specification.

Using the same international data sets, Waldinger explored the issue of whether tracking exacerbates the importance of family background for educational achievement and finds no such evidence once pre-tracking differences are controlled for. Waldinger claims that unobserved country-level factors may exist and affect the impact of the family background on students' test scores. He concludes that studies claiming the existence of a causal relationship between tracking and the importance of parental background need to be interpreted with caution.

Jenkins, Micklewright and Schnepf (2006) looked at levels of social segregation in England and compared the situation between English secondary schools with that in 27 OECD countries in addition to Scotland and Northern Ireland. They found that England is a middle-ranking country in terms of social segregation which mainly results from children from different social backgrounds being spread unevenly within the state sector. Parental choice was higher in English secondary schools and school choice was found to be low. However, they found that is was countries with a higher prevalence of school rather than parental choice that had higher levels of segregation.

A number of studies (Burgess, McConell, Propper and Wilson, 2005; Burgess, Propper and Wilson, 2005; West and Hind, 2006) have looked at the impact of choice, sorting and selection of students in selective and non-selective schools.

Burgess, McConell, Propper and Wilson (2005) investigated the impact of choice on pupils' sorting across schools using student-level data from England. They reached the following conclusions. a) After having established that most students in England are able to exercise choice between different schools, they then conclude that over half of the students in their sample do not choose the closest school to their place of residence. b) In the three dimensions of ability, socio-economic status and ethnicity, there is variation in the degree of student sorting across the country, but particularly among high ability students. Selective areas show the highest degree of sorting whereas they do not have high levels of neighbourhood sorting of ability. Areas with a bigger number of schools generally exercise more ability sorting.

Burgess, Propper and Wilson (2005) used evidence from economic research to explore the issue of whether more choice can improve outcomes in education and health care. They argue that the competitive pressure introduced by choice will improve standards. House prices will also fall which will be of particular benefit to working class families. More flexibility in the supply of

school places is needed to enable existing schools to expand, new schools to start and poor schools to close. It is argued that parental choice in addition to flexibility in the supply of school places will reduce 'sorting' of students by income and ability. School choice, however, can be affected by the role of peer groups if this largely influences parents' school choice.

West and Hind (2006) looked at the admissions criteria that are used by comprehensive secondary schools in London. Comparisons were made with the rest of England though these were restricted by the small school sample. The results indicated that those schools that are responsible for their own admissions have generally fewer pupils with special educational needs and perform better achieving 'higher positions in the published examination league tables'.

Manning and Pischke (2006) used the NCDS data source to show that studies using value-added measures may not succeed in removing selection bias between comprehensive and selective school students. The authors claim that any differential outcomes in performance at age 16 may not be necessarily due to the comprehensive or selective schooling system and, therefore, results need to be interpreted with caution. However, it needs to be reemphasised that studies using the 1958 birth cohort NCDS data are problematic regarding their relevance to today's context.

In spite of the valuable and rich information that the NCDS dataset provides, the age of the data limits its applicability to the current school situation as has been mentioned previously.

Levačić and Marsh (2007) explored the performance and resourcing of secondary modern schools in England and argued that these schools have larger numbers of socially and educationally disadvantaged pupils. Matching pupils' performance at KS2 and GCSE, they found that GCSE performance was slightly lower for pupils at secondary modern compared to comprehensive schools whereas grammar school pupils' attainment was considerably higher. Pupils of average ability gained most by attending grammar school. They then looked at the financing of secondary schools in England and concluded that secondary modern schools are underfunded in relation to the needs of their pupils resulting in these pupils being discriminated compared to grammar school pupils.

Maurin and McNally (2007) explored the effects on educational attainment of a reform which provided widened academic access to schools (more borderline pupils attending grammar schools) in Northern Ireland using England as the comparison group. Their findings indicated that the increase in the number of pupils entering the more academic track was followed by an increase in GCSEs and A-levels examination results. They also found that girls were more likely to enter grammar school during this 'open enrolment' period and this was followed by achievement gains arguing for gender-related differential educational outcomes from access into such schools. They also argued that the school system perpetuates any pre-existing inequality that may exist between socio-economic groups as the selective system was found to benefit pupils coming from a poor family background but access to such schools was unequal for this group of pupils.

Clark (2007a, 2007b) conducted a within-authority comparison to explore any differential effects of attending grammar schools as opposed to secondary modern schools on borderline pupils. Even though grammar schools were

not found to improve performance on test scores, they seemed to have large effects on course-taking, especially for pupils of lower socio-economic status. Grammar schools, therefore, offered these pupils more opportunities and choice in comparison to secondary modern schools. However, the age of the dataset (early 1970s) and its relevance to one authority only limit to a large extent the generalisability of the study as well as its applicability to the current educational context.

5.2. *Summary of the methodological criticisms of existing studies*

In Chapter 4 we identified a number of methodological issues that arise in trying to understand the impact of selection on attainment. Many of these are complex and technical. For now, we group them under five broad headings.

The first major, general methodological problem is that of controlling adequately for all the differences between pupils in different systems. For example, pupils at the borderline of grammar school acceptance who are accepted are likely to be different from those who are not, even if they have the same KS2 scores. But in most cases we do not have good information about these differences, so cannot take it into account in our comparisons. Even where we appear to have some measure of an important initial difference, in practice, less than perfect reliability of the measure creates a similar version of this problem.

The second problem is the quality of the data. All studies are limited by their data. Those that used the NCDS data were relatively well off in this respect, since a lot of good measures were available. These studies have a different, but no less important deficiency, however, which is the lack of representativeness of their samples (see fifth point, below). With the advent of matched pupil data from the national pupil dataset, researchers have been able to conduct analyses of the performance of all state school pupils in England. However, KS2 data (especially if it is limited to levels rather than raw marks) is not a perfect measure of prior attainment, and Free School Meals is a poor measure of socio-economic status. Outcome data are also less than ideal. GCSE grades may fail to discriminate sufficiently, especially at the top end; they may distort the relative sizes of the gaps between different levels of achievement. Aggregated measures of overall GCSE achievement may be sensitive to differences in the entry policies of different types of schools and the difficulty of different subjects. It would also be helpful to know something about performance after age 16, though matching any such data goes well beyond present availability; information about non-academic outcomes also seems rather limited.

The third major problem is in the methods of analysis that have been used. Different methods may well lead to different conclusions, even when there are no clear *a priori* grounds for preferring one method to another – or rather there are good reasons in favour of each.

Fourth is the difficulty of defining selective systems in order to make a comparison. This issue has received relatively little attention but seems highly problematic. Differences in the level and methods of selection are hard to take into account; movement across LA borders seems to have been generally underestimated in the desire to make a neat comparison. Furthermore, it is clear that 'comprehensive' systems are not wholly without their own forms

of selection. West and Hind (2006), for instance, found that those 'comprehensive' schools that are responsible for their own admissions have fewer pupils with special educational needs and perform better.

Our fifth and final difficulty relates to the samples used. Even those studies using national pupil datasets are not immune to this, since a proportion of pupils (almost certainly not a representative group) will be hard to match from KS2 to KS4, though the use of data from a complete national cohort is certainly a big advantage. However, the studies that used the 1958 birth cohort NCDS data are particularly problematic in this respect. The fact that these studies relate to the experiences of a group who went through secondary education at a time of significant change and over thirty years ago must raise questions about their relevance to today's context.

5.3. *Conclusions from the literature review*

Most studies that have been discussed in Chapter3 had focused on pupil and school outcomes when comparing selective and non-selective systems by concentrating on academic attainment (GCSE grades) and /or progress (value added measures). Different datasets as well as techniques of analysis have been employed in the various studies. There was no consistency in the results of the studies as to whether a selective or non-selective school system was of most benefit to its pupils. (e.g. Steedman, 1980, 1983; Jesson, 2000 & 2001; Galindo Rueda & Vignoles, 2004). Some studies found that the selective system benefits certain pupils (e.g. borderline students: Schagen and Schagen, 2003 & Clark, 2007a; Maurin and McNally, 2007; FSM pupil: Atkinson *et al*, 2004; average ability pupils: Levačić and Marsh (2007); and high ability girls: Galindo Rueda & Vignoles, 2004). Some results indicated that not having a grammar school available was detrimental to the achievement of highly able pupils (Atkinson *et al*, 2004). However, there have also been studies that claimed to have found results that support either one of the education systems. Jesson, (2000 & 2001) found pupils from comprehensive school pupils did better than pupils from selective school systems, but the DfEE data used in Jesson's studies have been highly criticised. There are other studies that have found greater benefits in the selective system as opposed to the comprehensive system (e.g. Marks *et al*, 1983 and 1985; and Sullivan and Heath, 2003). Such inconsistencies in the findings make it difficult to simply accept any one conclusion.

Many researchers have acknowledged the problem of comparing the effects of selective and comprehensive school systems as the existence of selective schools is piecemeal and creaming to some extent occurs to comprehensive schools which are in proximity to selective schools (Fogelman, 1984). Even when differences in the intake of different types of school can be controlled for adequately, there still remain considerable difficulties (Goldstein, 1984). Marks and Cox (1983) stated that the 'ideal' comparison would be between the comprehensive system as it is now with what the selective system would be now if Circular 10/65 had not been implemented, given all the other changes such as increasing resources, greater equalisation of resources between pupils in different schools, the raising of the school leaving age, the rise of CSE and the increasing proportion of girls succeeding in public examinations. This statement was made over 20 years ago, and it appears now that no such comparison will ever be possible.

There are several issues inherent in the difficulties of trying to find how various aspects of a school effect whole school or individual pupil achievement. These issues include defining particular types of school and the 'creaming' issue, suitability of input (explanatory) and outcome measurements, as well as the unit of analysis used (pupil / school/ LA) and inferences made from the results. Goldstein, highlighted these issues in 1984, and they are still apparent in the research in our literature review. Goldstein claimed (1984) that there were deficiencies in all existing comparative studies of school outcomes. As factors other than school types are also involved in influencing pupils' performance, Goldstein suggested that studies comparing pupils' attainment between schools should have as much information about the pupils, schools and LEAs and utilise a multistage and multilevel study design. However, multilevel modelling does not seem to have resolved these issues.

More recent studies have employed the multistage and multilevel design (e.g. Yang & Woodhouse, 2001; Schagen & Schagen, 2003; and Levačić and Marsh, 2007), but the data used were limited. Yang & Woodhouse investigated pupils' progress by focusing on pupils with A-levels and gathering information on their GCSE examinations, and did not have information on pupils' socio-economic characteristics, mobility or social environment of the establishments. Although Schagen & Schagen's and Levačić and Marsh's analyses controlled for social class, eligibility for Free School Meals as a suitable variable to measure social class should be questioned (see discussion of this issue in Section 4.2.4, p112).

The NCDS data is perhaps the most comprehensive dataset to date as it includes test scores, behaviour measurements, exam results as well as information about pupils, schools and LEAs. The data have been used in more recent studies (Sullivan & Heath, 2003; Galindo-Rueda & Vignoles, 2004) to compare the effects of the different education systems, but ended up with somewhat different conclusions. Steedman (1980 and 1983) found that pupils' exam attainments were not systematically different in a way that permit the conclusion that going to comprehensive schools as opposed to a system of grammar and secondary modern schools explained examination performance. Similar to Steedman, Galindo-Rueda and Vignoles' (2004) found that a selective system benefits some pupils but they could not find a systematic significant effect, either positive or negative. Sullivan and Heath (2003) however concluded that pupils in selective schools achieved superior educational outcomes to pupils in comprehensive schools.

Sullivan and Heath included pupils from independent schools in their analysis and they analysed pupils' attainment independently based on the different types of schools. Although it is possible that the results achieved by pupils in secondary modern schools lower the composite score for the selective system as compared to the comprehensive system, it is notable that Sullivan and Heath found that pupils' performance in the secondary modern schools was better than pupils in the comprehensives across all their outcome measures. It is likely that the techniques of analysis together with a relatively small representative sample from the different types of schools lead to the different findings and conclusions from The NCDS data. Nevertheless, it is not appropriate to apply the findings based on the NCDS data to today's situation, as over 30 years later, the sample used does not represent the current situation of organisation of schools in England

Instead of analysing raw examination results, some researchers have analysed students' progress, or value added, as a measure of effectiveness of the different education systems. This method allows for variation in student intake by measuring the achievement of each pupil relative to his or her starting point, but caution should be adopted when calculating and interpreting a value added measure. The DfEE method of calculating value added has been criticised on several points, particularly it being insensitive to the 'ceiling' effect. Other value added measures have been used in more recent research. Manning and Pischke (2006), however, emphasised that such studies are unlikely to be entirely successful in eliminating selection bias between comprehensive and selective school students and, therefore, any differential outcomes in performance at age 16 may not be necessarily due to the comprehensive or selective schooling system.

Comparing selective and non-selective systems is beset with the problem of comparing like with like. In the two studies that did attempt this through propensity matching, pupils were matched in terms of their social background and/or social environment, but variables that may influence school intake such as characteristics of the catchment area and proximity of selective and comprehensive schools were not matched. Many argue that the existence of selective schools is creating an imbalance to the intake of comprehensive schools that are in proximity to them, and thus they are not truly comprehensive schools. The existence of independent schools is likely to have a similar effect. Croxford (2000) suggested that the existence of independent schools in a way creates an imbalance to the ability level of children in comprehensive schools, but we haven't seen any studies that have considered this.

5.4. Discussion

What is clear from our attempts to evaluate the findings from existing studies is that there is no clear consensus; even studies using the same data do not seem to agree about the results. Across different datasets, times, methodologies, and of course researchers, there appears at first sight to be hardly any agreement.

However, disagreement in such cases can sometimes be more apparent than real. Results which are really quite similar and certainly perfectly commensurable with each other can be presented quite differently and may even appear contradictory. The best way to create an overview of such disparate findings is by the process of meta-analysis. This technique allows an overall average effect to be calculated across all studies, suitably weighted, for the homogeneity of apparently different effects to be tested, and, if heterogeneous, for factors that may explain different effects to be investigated.

Unfortunately, however, any attempt to combine the results of different studies to produce an overall answer seems to be impossible at this stage. We had certainly hoped to be able to conduct some kind of quantitative synthesis, or meta-analysis, of the results to provide an estimate of the overall grammar school effect. Certainly, the variety of different methods and data types would have made this quite challenging. However, the real problem is the widespread and varied methodological weaknesses of the studies. If the results of every study are threatened by serious problems such as

methodology, data and sampling, then there seems little point in calculating an average; an average is only as good as its constituent parts.

Given this failure to conduct a quantitative synthesis, we are left with a rather impressionistic summary, though a number of points of agreement do appear to emerge.

Certainly, the majority of studies seem to find that pupils who attend grammar schools do better than equally able pupils in comprehensives. The only studies that positively disagree with this claim are those by Jesson (2000; 2001) and House of Lords (2000). If we limit our consideration to the methodologically stronger studies of 11-16 education in England, then this majority in favour of grammar schools becomes unanimous. This is true both for those that used national datasets (Schagen and Schagen, 2003, 2005; Atkinson *et al* 2004; Levačić and Marsh, 2007) and for those based on the NCDS data (Sullivan and Heath, 2002; Galindo-Rueda and Vignoles, 2004). Hence this looks like a fairly robust finding from the existing literature, supported by a strong consensus.

However, even with unanimous support from these stronger studies, we cannot rule out the possibility that this result could arise, at least in part, as an artefact of 'regression to the mean', since none of these studies makes any adjustment for it. The result also seems likely to be sensitive to the inclusion of school composition variables; very few studies have included these (only Schagen and Schagen, 2003, 2005) and none seems to have explicitly considered how they might affect the outcome. We return to these questions in Chapter 8 when we conduct our own analysis of a national dataset.

More equivocal in the literature is the matter of whether selective systems as a whole (i.e. grammar plus secondary modern) perform better than comprehensive systems. Of those who consider this question, some find no appreciable difference (Steedman, 1980; Atkinson *et al*, 2004), others find in favour of selective systems (Marks *et al*, 1983, 1985), while others favour comprehensive systems (Jesson, 2000, 2001). On this question, therefore, the literature is unclear.

Consensus seems even harder to reach on the further question of precisely which subgroups (if any) benefit from grammar school attendance. Each study seems to have its own particular group, with, for example, Schagen and Schagen (2003; 2005) Maurin and McNally (2007) and Clark (2007) arguing for those at the borderline of grammar school entry, Atkinson *et al* (2004) and Maurin and McNally (2007) finding it is particularly those eligible for Free School Meals, and Galindo-Rueda and Vignoles (2004) identifying able girls as the biggest beneficiaries. A particularly interesting finding that needs to be emphasised here concerns the benefits that pupils from lower socio-economic groups seem to gain from attending grammar schools (Jesson, 2007, Maurin & McNally, 2007, Clark, 2007) but the restricted access of these pupils to such schools perpetuating, according to Jesson, a cycle of deprivation and of limited opportunities. Interpreting these claims is particularly problematic as we do not know how many comparisons may have been tested in each study; if you do enough comparisons, you will eventually find one that shows a difference. Given the lack of consensus across the claims about which subgroups benefit, we have to conclude that again the literature is unclear on this matter.

Let us therefore try to summarize what is known about the impact of selection on performance. We have devoted more than 100 pages of this report to presenting and discussing the existing studies and the methodological issues they raise. There is a strong consensus that those who attend grammar schools do better as a result, though even here we have voiced doubts about the robustness of this finding. The question of whether selective systems as a whole are better or worse than comprehensive systems remains unresolved, as does the question of which particular subgroups (if any) benefit most from grammar school attendance.

PART III
ANALYSIS OF NATIONAL PUPIL DATASETS

6. Differences between selective and non-selective schools

In this chapter we first present a description of some of the differences between selective and non-selective schools and their pupils. We then present some analyses of where pupils live who attend grammar schools. Next, we consider the extent to which Local Authorities (LAs) can be identified as 'selective' or not. This leads to an examination of the extent to which pupils cross LA boundaries and how this differs for different types of school. Finally, we examine the types of qualifications taken by pupils in grammar schools and others and address the issue of subject difficulty.

6.1. Descriptive statistics

There are 164 grammar schools in England, just 5.2% of the 3,149 maintained, mainstream secondary schools in the country. In terms of pupil numbers, of the 592,475 children in Year 11 in these schools, 22,767 or 3.8% attended a grammar school in 2005/06. These are small proportions of the national picture, though their political importance is perhaps out of proportion to their size.

6.1.1. Pupil-level differences

Table 27 shows some descriptive statistics for pupils in grammar and non-grammar schools. If we look at the statistics for England the most striking difference is in the percentage of pupils eligible for free school meals (FSM), with proportionally more than six times as many pupils from these poorer homes in non-grammar schools (13.3%) as in grammar schools (2.0%). Although this is a large difference, it does not necessarily imply that grammar schools are unfairly privileging relatively advantaged pupils in their selection processes. Socio-economic status and ability are generally correlated, so it would be hard to select by ability without selecting a higher proportion of wealthy pupils. We return to this question in the next section. With regard to ethnicity, grammar schools appear to be slightly more mixed than other schools, though the difference is very small.

Table 27: Statistics for pupils in grammar and other schools – pupil level

	Grammar schools	Non-grammar schools
Total number of pupils (Year 11)	22,767	569,708
Percentage of all pupils (Year 11)	3.8%	96.2%
Percentage eligible for free school meals	2.0%	13.3%
Percentage 'White British'	80.0%	80.6%

6.1.2. *School-level differences*

Table 28 shows school-level descriptive statistics for grammar and non-grammar schools. There are several striking differences between the two. Firstly, grammar schools are far more likely to be single sex schools, in fact proportionally more than seven times more likely. Fewer than 10% of non-grammars are single sex schools compared to three quarters of grammars. Secondly, grammar schools are almost wholly made up of 11–18 schools compared to about half of non-grammar schools. The three grammar schools with an age range other than 11-18 have age ranges of 12-18 and 13-18 showing that all grammar schools have sixth forms. Smaller differences can be seen in the specialist status and school type figures. Proportionally more grammar schools are specialist schools, by about 12%. As would be expected from the historical background of grammar and non-grammar schools (see Section 1.1.1, p3) the breakdown of school type differs, with grammar schools being predominantly Foundation schools and non-grammars being predominantly Community schools.

Table 28: Statistics for pupils in grammar and other schools – school level

		Grammar schools (total = 164)		Non-grammar schools (total = 2985)	
		Number of schools	Percentage	Number of schools	Percentage
Gender of entry (Years 7 - 11)	Boys only	60	36.6	122	4.1
	Girls only	62	37.8	164	5.5
	Total single sex	122	74.4	286	9.6
	Mixed	42	25.6	2699	90.4
Age range	11-16	0	0.0	1294	43.4
	11-18	161	98.2	1464	49.0
	Other	3	1.8	227	7.6
Specialist status	Specialist	151	92.1	2383	79.8
	Non-specialist	13	7.9	602	20.2
School type	Academy	0	0.0	33	1.1
	City Tech. College	0	0.0	10	0.3
	Community	39	23.8	1920	64.3
	Foundation	80	48.8	442	14.8
	Voluntary Aided	33	20.1	497	16.6
	Voluntary Controlled	12	7.3	83	2.8

6.2. *Grammar school attendance in different parts of the country*

It might seem reasonable to suppose, and indeed many previous studies have done, that in selective LAs, or in the selective parts of partially selective LAs, pupils living there would attend a selective school (if they were selected), and that those living in other areas would not. In line with this prevailing assumption, our initial aim in conducting this analysis was to try to identify a set of 'selective areas' in which grammar school pupils lived and in which non-selective schools would be 'creamed' by those grammar schools. We anticipated that we could then compare the performance of pupils in these

'selective areas' with the performance of those in areas unaffected by selection.

However, when we came to look at the PLASC (Pupil Level Annual School Census – the DfES statutory school census) and Performance Table data, we discovered that, in reality, pupils were travelling large distances to attend grammar schools; a small but perhaps significant number were travelling extremely large distances; they were certainly crossing LA boundaries in large numbers. Hence the notion of a 'selective area' seemed hard to define; there were certainly no clear boundaries around such areas.

In this section, we present data showing where grammar school pupils live, analysed both by ward and by LA of residence, as well as by LA of school attended. We conclude that the LA is not a suitable unit of analysis for understanding the impact of selection.

6.2.1. Grammar school attendance by ward of residence

Figure 3 (p141) shows the distribution of grammar school attendance across England. Based on the pupils' home addresses, the percentage of pupils in each ward attending grammar schools has been calculated. Wards have then been coloured accordingly on the map, with the increasing colour depth corresponding to increasing percentages of pupils attending grammar schools. Areas where there is no grammar school attendance are not shaded in, but left white.

Care needs to be taken when interpreting this map as areas of equal shading equate to equal *proportions* of pupils attending grammar schools rather than equal numbers, so rural areas may appear to be impacted by grammar schools more than urban areas. However, with this in mind, it is still surprising to see such a high proportion of the country shaded in. This means that high proportions of the country have pupils attending grammar schools, even though there are only 164 grammar schools. It is particularly interesting to note the number of 'non-selective' LAs that show a high percentage of pupils attending grammar schools. For example there are wards in Dorset and Nottinghamshire (where there are no grammar schools) that are coloured dark red, indicating that there are many pupils crossing LA boundaries to attend grammar schools. It is also interesting to note the high proportion of areas of the darker colour red (indicating a high proportion of grammar school attendance) in rural Cumbria where there is only one grammar school. (See Table 29, p142 for comparison between grammar school attendance by pupil home LA and percentage grammar schools in each LA.). For England as a whole, of the 8,033 electoral wards for which we have data, a third (2,678, or 33%) have pupils living in them who attend a grammar school. However, it is also interesting to note, that this proportion has reduced considerably since 2004, when 47% of wards had pupils living in them who attended grammar schools, a drop of 14%.

6.2.2. Grammar school attendance by LA of residence

Table 29 shows the percentage of pupils living in each LA area (i.e. local authority or county council area) who attend grammar schools (limited to LAs where more than 1% attend grammar schools). The table also shows the proportion of that LA's secondary schools that are selective.

Table 29 shows similar patterns to the maps in that pupils from over 40 'non-selective' LAs cross boundaries in order to attend grammar schools. For example 11% of pupils in Windsor and Maidenhead attend grammar schools despite the fact that the LA has no grammar schools. It is also interesting to note that 29% of schools in Reading are grammar schools, whereas only 4% of pupils in Reading attend grammar schools.

Figure 3: Map of percentage grammar school attendance by ward of residence (England)

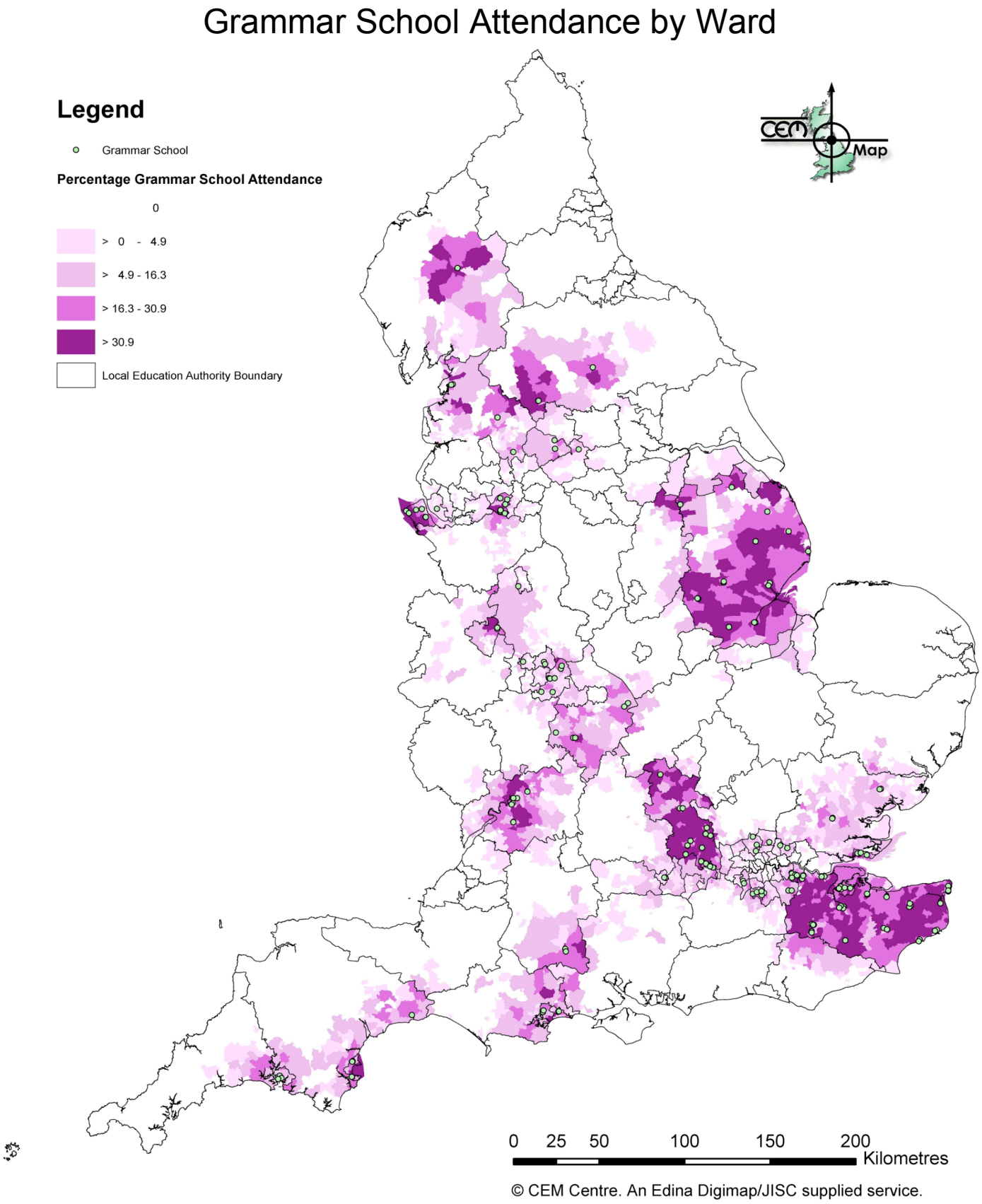


Table 29: Percentage grammar attendance by LA of pupil home address, compared with percentage grammar schools in LA

Local Education Authority	% pupils attending grammar schools	% grammar schools
Buckinghamshire	34%	38%
Trafford	32%	39%
Kent	28%	32%
Medway	27%	32%
Slough	25%	36%
Bexley	24%	25%
Wirral	22%	27%
Lincolnshire	22%	24%
Southend on Sea	19%	33%
Torbay	18%	38%
Poole	15%	25%
Bournemouth	15%	20%
Sutton	14%	36%
Kingston upon Thames	13%	20%
Gloucestershire	12%	17%
Calderdale	11%	13%
Windsor and Maidenhead	11%	0%
Plymouth	10%	18%
Bromley	9%	12%
Greenwich	7%	0%
Barnet	7%	15%
Merton	6%	0%
Redbridge	6%	12%
Birmingham	6%	11%
Warwickshire	6%	14%
Croydon	5%	0%
Walsall	5%	10%
Wokingham	5%	0%
Enfield	4%	6%
Richmond upon Thames	4%	0%
North Yorkshire	4%	7%
Dorset	4%	0%
Wiltshire	4%	7%
Reading	4%	29%
Devon	4%	3%
Essex	4%	5%
Telford and Wrekin	4%	14%
Harrow	3%	0%
Hillingdon	3%	0%
Sandwell	3%	0%
Solihull	3%	0%

Local Education Authority	% pupils attending grammar schools	% grammar schools
Wolverhampton	3%	6%
Kirklees	3%	4%
North East Lincolnshire	3%	0%
Milton Keynes	3%	0%
Lancashire	3%	5%
Lewisham	2%	0%
Brent	2%	0%
Haringey	2%	0%
Hounslow	2%	0%
Liverpool	2%	3%
Manchester	2%	0%
East Sussex	2%	0%
Staffordshire	2%	0%
Bracknell Forest	2%	0%
West Berkshire	2%	0%
Cumbria	2%	2%
Surrey	2%	0%
Hackney	1%	0%
Islington	1%	0%
Lambeth	1%	0%
Wandsworth	1%	0%
Ealing	1%	0%
Waltham Forest	1%	0%
Dudley	1%	0%
Knowsley	1%	0%
Bury	1%	0%
Rochdale	1%	0%
Salford	1%	0%
Stockport	1%	0%
Bradford	1%	0%
North Lincolnshire	1%	0%
Rutland	1%	0%
Stoke-on-Trent	1%	6%
Cheshire	1%	0%
Warrington	1%	0%
Worcestershire	1%	0%
Nottinghamshire	1%	0%
Shropshire	1%	0%
Cornwall	1%	0%
Hertfordshire	1%	0%
Oxfordshire	1%	0%

6.3. *Crossing LA borders*

6.3.1. Border crossing, by LA of school attended

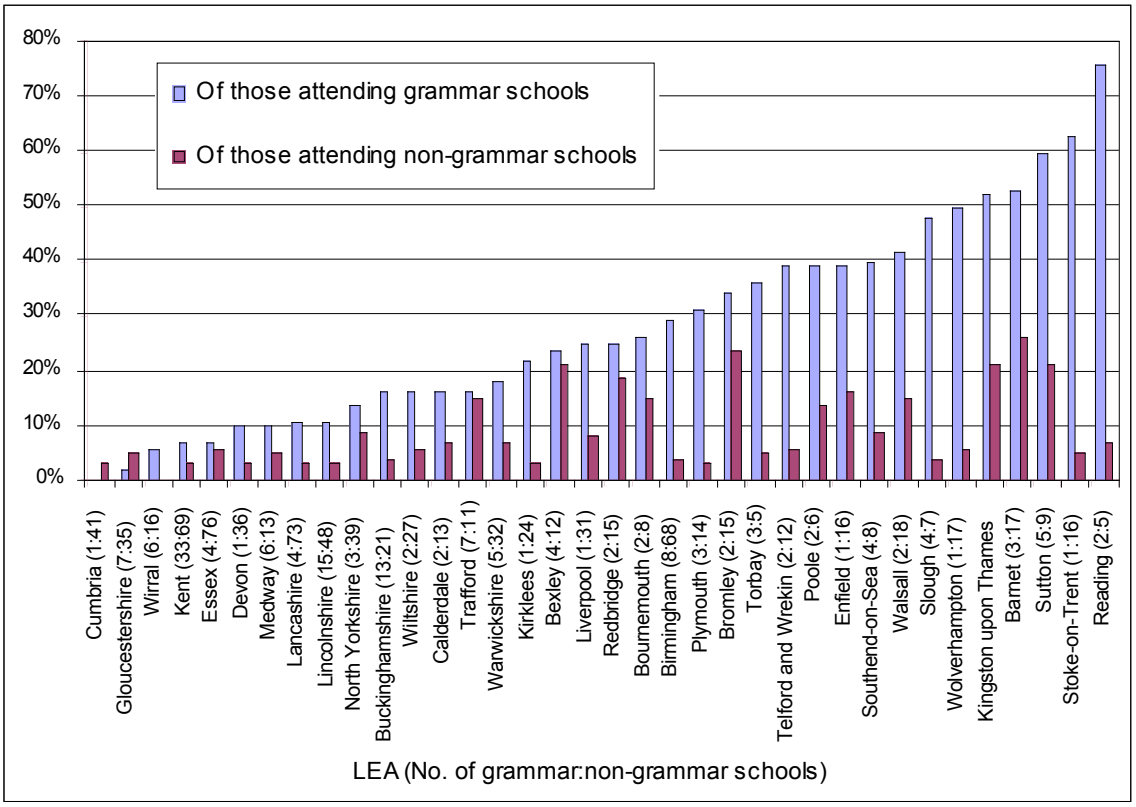
Figure 4 shows the proportion of pupils crossing LA boundaries to attend school, split by the LA of the school attended. Only those LAs that contain at least one grammar school are listed. The results are further split by those attending grammar schools and those attending non-grammar schools. The

numbers of grammar schools and non-grammar schools in each LA are also shown.

Firstly we note that the percentage crossing LA boundaries varies considerably by LA, with the majority of students in Reading crossing boundaries to go to school, compared to Cumbria, where boundary crossing is rare (which is probably not surprising given the geography).

The second thing to note is that in all LAs the proportion of pupils crossing boundaries and attending grammar schools is greater than the proportion who cross and attend non-grammars. However, the difference in these proportions varies considerably, with almost all boundary crossing from Reading being due to pupils attending grammar schools as compared to Bexley and Essex where the split between grammar and non-grammar schools is more even. It is likely that the LA size and the urban/rural nature of the LA are playing a part in this movement.

Figure 4: Percentage of pupils crossing LA boundaries to school, by LA of school attended



6.3.2. *Is the LA a suitable unit of analysis?*

Analysis of the 2004 PLASC data shows that, for pupils in England as a whole, about 8% cross an LA boundary to go to secondary school. Of those who attend a grammar school, the corresponding percentage is 20%. This is a substantial difference, and indicates that, perhaps not surprisingly, those who attend grammar schools tend to travel further to school. However, the fact that overall, one in five of the pupils in grammar schools have come from outside the LA casts considerable doubt over the validity of any analysis that assumes we can use the LA as a way of identifying areas where there is or is not an effect of selection. Moreover, as shown in Figure 4, in a large number

of 'selective LAs' the proportion of grammar school pupils crossing the boundary is well above one fifth; in several it is more than half. Although there is some correlation between the proportion of pupils living in an LA who attend grammar schools and the proportion of that LA's provision comprising such schools (Table 29), there are also some large differences.

This issue was raised in Chapter 1 and the reader is referred back to the discussion in that section. The results of the analysis reported here seem to be conclusive: the LA is not a suitable unit for analysing selective and non-selective systems. From this, two clear implications follow.

The first is that our results confirm that the studies analysed in Chapter 3 that have used such a unit (e.g. Jesson, 2000, 2001; Schagen and Schagen 2003) are somewhat problematic, as indeed was suggested in Section 4.5 (p116).

The second implication is that we must now ask, 'Is there some other way we can define selective units of schools or areas, if not the LA?' We return to this question in Chapter 7 when considering the extent to which non-selective schools are 'creamed' by selective schools.

6.3.3. Border crossing, by FSM status

One further question we can ask, in relation to crossing LA borders to attend school, is whether there are any differences between pupils who are eligible for free school meals (FSM) and those who are not. For example, if it were the case that those who crossed a border to attend a grammar school were disproportionately from more advantaged backgrounds (i.e. non-FSM), then this might indicate that for poorer children, financial or other barriers to travel were unfairly restricting their access to grammar schools.

Taking all schools together, FSM pupils are less likely to have crossed LA boundaries (5.4%) compared with those not eligible for FSM (9.1%). For those at grammar schools, the percentages crossing borders are 17.8% and 19.9% for FSM and non-FSM respectively. This difference is bigger for pupils all schools together than it is for those in grammar schools and if we compare the odds ratios (0.57 and 0.87 respectively), the difference is enough to be statistically significant. This indicates that FSM is more of a factor in crossing LA boundaries for non-grammar school pupils than for grammar school pupils – non-grammar school pupils are more likely to travel further to a non0grammar school, but it makes little difference for grammar school pupils. The same calculations were performed for 2004 data and the opposite was found to be true, but the results were not statistically significant.

6.4. Subject difficulties and exam-entry profiles

In this section we attempt to answer two questions: are KS4 subjects of equal difficulty and if not, does the exam-entry profile of grammar schools differ from non-grammar schools?

6.4.1. Why consider subject difficulty?

The Government points systems for Key Stage 4 qualifications places qualifications such as GCSEs (including short courses and vocational courses) on the same points scale. Equal points for differing subjects and qualifications assumes equal difficulty and the assumption follows through to league tables and entry requirements for further and higher education. It is therefore

important to find out whether this is the case. If it is not then caution needs to be taken when using point scores to compare achievement, whether between individuals or institutions.

6.4.2. Method of calculating subject difficulties

The methodology for this section has been taken from Coe, 2008. In this paper Coe establishes subject difficulties by analysing the KS4 2004 data from the National Pupil Database using Rasch analysis. The Rasch model (Rasch, 1960/1980; Wright & Stone, 1979) provides a method for calibrating ordinal data onto a scale that is adequate for measurement, with properties such as unidimensionality, linearity, sample and scale independence (Wright, 1997). Unlike other statistical models, Rasch turns the relationship between data and the model upside down. Whereas most statistical modelling attempts to fit a model to existing data, in Rasch the model comes first, since the model embodies the precise requirements for adequate measurement. If data do not fit the model we must reject the data, not the model.

Rasch assumes that the 'difficulty' of items and the 'ability' of persons can be measured on the same scale. Rasch analysis uses an iterative procedure to estimate item difficulties and person abilities for a given data set. It allows the fit of the model to be investigated and misfitting items and persons to be identified.

In the context of GCSE examination data, each subject may be thought of as an 'item', although each subject has a number of levels of success (grades). Hence a partial credit model can be used, in which the difficulty of each grade within each subject is estimated separately. The current analysis was conducted using WINSTEPS (Linacre, 2005a).

The partial credit model treats each grade as defining a threshold between those who have achieved that grade (or higher) and those who have achieved a lower grade. Hence it does require an assumption about the order of grades within a subject (e.g. that A is higher than B) but makes no assumptions about the relative sizes of the gaps between them, or about the equivalence of the 'same' grade in different subjects.

The estimate of the difficulty of a particular grade in a particular subject is based on all the candidates who have taken that subject with at least one other. The grade difficulty depends on the relative probabilities of that grade being achieved by candidates of different ability, as determined by their performance in all their subjects and taking into account the different difficulties of all the grades they have gained.

The analysis in Coe (2008) has been repeated for this study using 2006 data from the National Pupil Database. Of the 592,475 cases in the dataset, 20,766 had achieved no pass grade (G-A*) in any of the 59 GCSE subjects available. Excluding arts (art and design, music and fine art) and minority language subjects (e.g. Danish, Arabic, Urdu, Persian, etc) from the analysis in order to get the Rasch model to fit meant that a further 1,263 students were excluded. The 36 subjects included were:

English literature, double science, design technology-electronics, design technology-food, design technology-graphics, design technology-resistant materials, systems, design technology-textiles, history, geography, French, German, business, RS, short RE, PE, physics, chemistry, biology, drama, IT, short IT, Spanish, maths, English, single science, statistics, media, office

technology, home economics-child development, social science citizenship, vocational business, vocational health and social care, vocational leisure and tourism, vocational science, vocational IT.

The model achieved person separation reliability of 0.94. INFIT and OUTFIT values for all subjects were between 0.70 and 1.71. For every subject the correlation between Rasch estimates of students' abilities and the grades achieved in that subject ('point-measure correlation') was at least 0.75, with the mean across all 36 subjects of 0.82. Principal Components Analysis of Rasch residuals showed that 81% of the observed variance was explained by the measures, with no more than 1% explained by any residual contrast. Overall these figures indicate that the 36 subjects fit reasonably well to the model. The model provides a measure of each student's overall academic achievement that is highly reliable and uni-dimensional.

In order to use this Rasch measure as an outcome in estimating grammar school effects we scaled the original measure (in logit units) into GCSE grade units. The Rasch measure of overall achievement at GCSE (which takes into account the relative difficulties of different grades in different subjects) was scaled to have the same mean and standard deviation as the average GCSE score of the same students. Hence a difference of 6 points on this scale corresponds to one GCSE grade on average (see Table 30 for the relative point scores by grade and qualification type).

Table 30: Relative point scores by grade and qualification type

Qualification type	Relative weighting	Grade	Qualification point score
GCSE	1	A*	58
		A	52
		B	46
		C	40
		D	34
		E	28
		F	22
		G	16
Short course GCSE	0.5	A*	29
		A	26
		B	23
		C	20
		D	17
		E	14
		F	11
		G	8
Vocational GCSE	2	A*	116
		A	104
		B	92
		C	80
		D	68
		E	56
		F	44
		G	32

6.4.3. Results

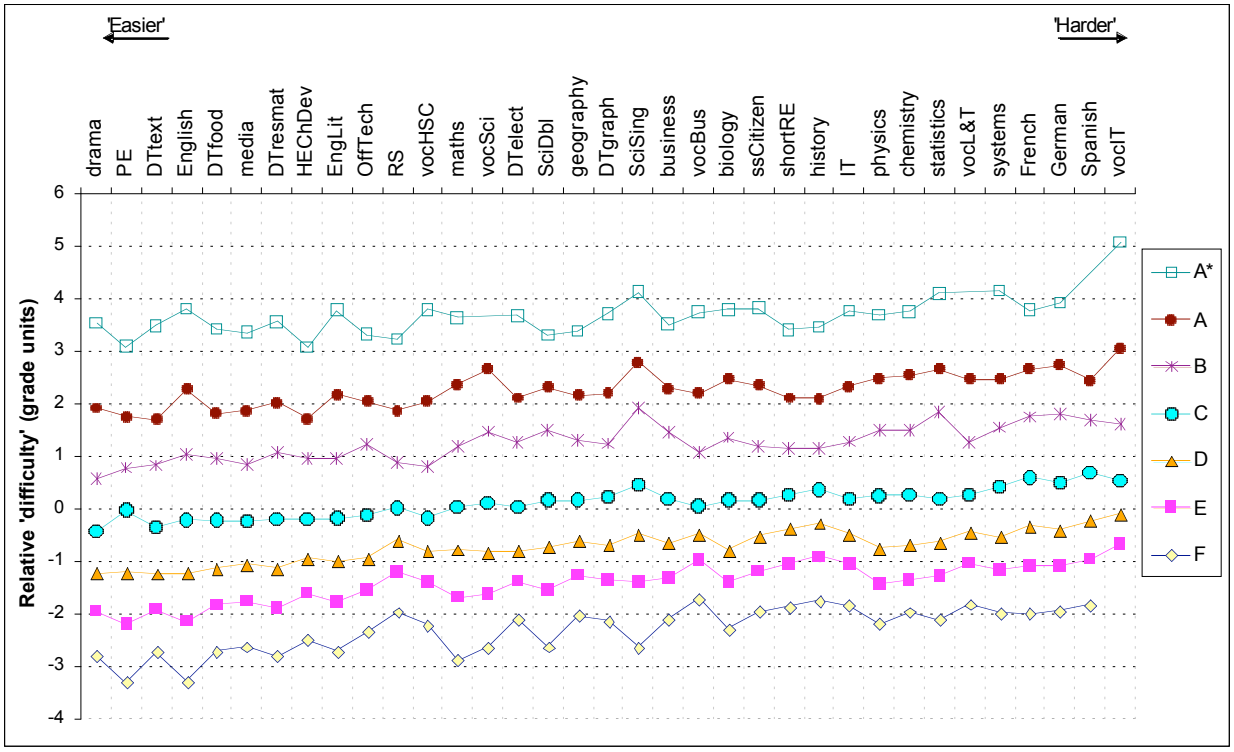
Table 31 shows the results of applying the Rasch technique to the NPD data. Difficulties have been expressed in terms of correction factors needing to be added to the scaled point scores to make all subjects equivalent. Subjects have been ordered in ascending order of difficulty. The results show that subjects vary considerably in difficulty. There is a difference of nearly two grades between the easiest subject (physical education, -0.84) and the hardest subject (vocational IT, 1.01) i.e. an A grade in GCSE physical education is equivalent to a B/C grade in vocational GCSE IT.

Figure 5 presents the relative 'difficulty' of achieving each grade. It is clear from this chart that intervals between grades tend to be bigger at the top than at the bottom.

Table 31: GCSE Difficulties (in grades) – Increasing from easiest through to harder subjects

Subject	Overall difficulty (in grades)
physical education	-0.84
drama	-0.68
design technology - textiles	-0.66
English	-0.56
design technology - food	-0.55
media, film & TV studies	-0.53
home economics - child development	-0.51
design technology - resistance of materials	-0.49
English literature	-0.36
office technology	-0.28
maths	-0.24
vocational I health & social care	-0.20
religious studies	-0.17
double science	-0.13
design technology - electronics	-0.05
geography	0.01
design technology - graphics	0.03
biology	0.05
vocational science	0.06
business studies	0.07
physics	0.10
short religious studies	0.11
social science - citizenship	0.16
vocational business	0.17
chemistry	0.20
history	0.22
IT	0.23
single science	0.34
statistics	0.34
systems	0.37
vocational leisure and tourism	0.40
Spanish	0.44
French	0.46
German	0.50
short IT	0.99
vocational IT	1.01

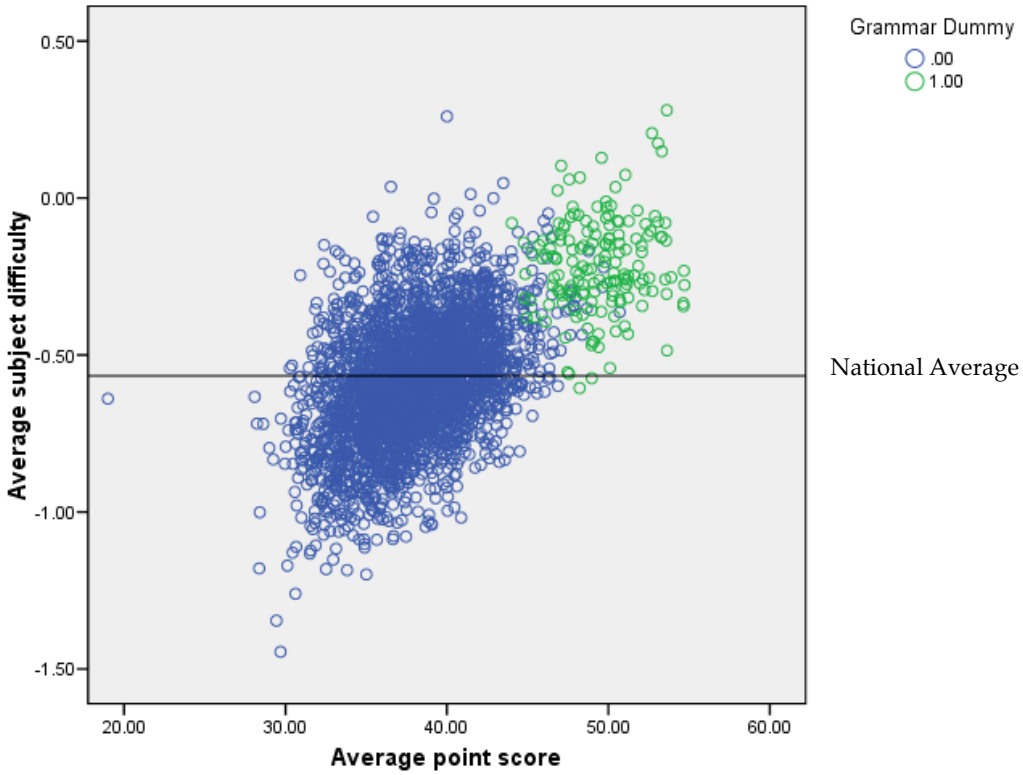
Figure 5: Relative 'difficulties' of achieving each grade, ordered by weighted average difficulty



6.4.4. *Comparison of KS4 entry profiles for different school types –school level*

With the subject difficulties calculated, we can now look for relationships between subject difficulty and schools/pupils. Only the GCSEs for which we have subject difficulties were used in the analysis. At pupil level *average difficulty* is used as a measure of the difficulty of the subjects passed by each pupil. Before averaging, the difficulty (in grades) of each subject taken by each student was converted into points (e.g. by multiplying a single GCSE difficulty by six). The resulting average difficulties are therefore measured in points. At school level average subject difficulty is calculated by taking the mean of all grades achieved by pupils in the schools (again this is presented in terms of points). This enables us to look at the relationship between difficulty and achievement by plotting average difficulty against average point score. We begin with the school level data:

Figure 6: Comparison of average subject difficulties for grammar and non-grammar schools



The horizontal reference line on Figure 6 represents the average subject difficulty for schools (note that in the case of Rasch scales zero does not represent average difficulty). From the chart we can clearly see that grammar schools are fundamentally different to non-grammar schools. Not only do the grammar schools differ in their higher average achievement (which we would expect), but they also differ significantly in the average difficulty of subjects taken. Almost all grammar schools appear above the reference line, showing that in the majority of grammar schools, pupils are choosing (or being offered) the more difficult subjects.

6.4.5. Comparison of KS4 entry profiles for different school types –pupil level

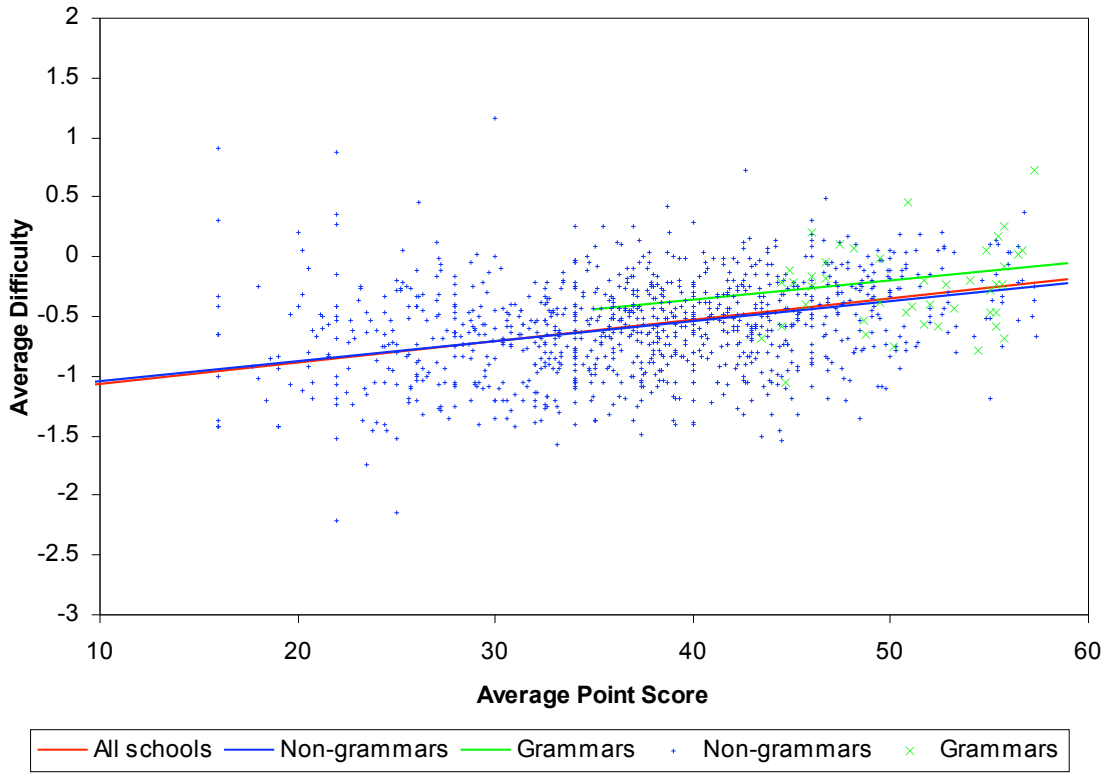
We have seen that pupils in grammar schools are sitting more difficult subjects than pupils in non-grammar schools. However, it could simply be that the more-able pupils, in whatever school, tend to be entered for harder subjects. Perhaps the more-able pupils in non-grammar schools are behaving no differently to similar pupils in grammar schools and the difference we see is due to grammar schools selecting the more academically able pupils, rather than to any differences in curriculum or entry policies once they are there. To resolve this issue, we need to compare like with like at the pupil level to ascertain whether after taking average points score into account, pupils at grammar schools are still taking the harder subjects (See Figure 7).

In order to compare subject difficulty between grammar and non-grammar schools we have calculated regression lines for average difficulty (at the pupil level) against the pupils' point scores. If pupils in grammar schools are taking more difficult subjects simply because they are the more able pupils, we would expect the more able pupils in other schools to follow the same

pattern. In which case we would expect the regression lines for grammar and non-grammar schools to be identical.

These regression lines have been calculated based on *all* the pupils, however only a random sample of 10000 pupils are shown as dots and crosses on each chart to give the general idea of the pattern of the results.

Figure 7: Comparison of average subject difficulties for pupils at grammar and non-grammar schools



Firstly we can clearly see that subject difficulty increases with academic ability in both grammar and non-grammar schools. However there is a clear difference between the regression lines of the grammar and non-grammar schools showing that academically more able pupils at grammar schools take even more difficult subjects than similar pupils at non-grammar schools. In fact a typical grammar school student with an average point score of 50 is choosing subjects 0.2 of a point harder than similar pupils in non-grammar schools. Although this appears quite small, across a year group of 150 pupils, each doing 10 subjects, this would add up to around 50 slipped grades. In other words if these 150 grammar school students had chosen subjects more typical of non-grammar school students 50 of their 1500 grades would be better.

6.4.6. *Implications for comparisons of achievement*

We have shown that there is a modest difference between the difficulty of the subject mix taken by pupils in grammar schools and those in non-grammar schools. Hence it may be important to make a correction for subject difficulty when comparing the performance of pupils in the two types of school.

7. Relationships between selective and non-selective schools

In this chapter we describe and analyse two concepts – *creaming* and *selectivity*.

The 'creaming' effect is based on the idea that the existence of grammar schools means that pupils who attend them are in effect taken from other schools in the same area. We use this concept to try to quantify the effects of grammar school selection on other schools and to try to identify groups of schools that function as a 'selective system'.

The notion of 'selectivity' is concerned with the differences between the pupils who attend a particular school and their neighbours who attend other schools. In particular, we are interested in the question of whether grammar schools are more or less socially selective than we might expect from their level of academic selection. We also investigate the different levels of 'selectivity' of selective and non-selective schools.

7.1. Creaming

Having shown that the LA was not a suitable unit for analysing the impact of selection, we were left with the idea that we might be able to identify groups of schools that behaved as a 'selective system'. Such a group would consist of one or more selective schools, together with a number of non-selective schools that were 'creamed' by them. Such selective systems of schools would not necessarily follow LA boundaries, nor would they necessarily respect the official classification of non-selective schools as 'comprehensive' or 'secondary modern'. Schools in 'non-selective' LAs (in the sense that the LA had no selective schools of its own) could well turn out to be heavily creamed, and schools that were creamed by grammar schools would not necessarily be designated 'secondary modern', but could be just as likely to be 'comprehensive'. Nor did it seem to be necessarily the case that selective schools would cream pupils from their immediate neighbours; there might be far more complex patterns of movement at work.

Hence, there seemed to be a need to define the relationship between selective and non-selective schools in terms of their impact on pupils, rather than in terms of the official designation or geographical or administrative location. We describe first our attempts to do this, and then how we tried to identify selective systems of schools, based on this relationship.

7.1.1. Calculating the 'creaming' effect of any given grammar school on any given non-grammar school.

In order to do this, we developed a procedure for calculating the 'creaming' effect of any given grammar school on any given non-grammar school. This

process is quite complicated, but needs to be understood in order to appreciate the findings.

We first present a justification of the method. Second, we give a simplified example to illustrate the process. Third, we present a detailed description of the steps in the procedure. In order to simplify the illustration, we assume that there are only two kinds of school: grammar schools (selective) and comprehensives (non-selective).

7.1.2. Explanation and justification for the method of identifying 'creaming'

The method rests on the idea that if a comprehensive school recruits pupils from a particular ward and some of the other pupils in that ward go to a grammar school, then that comprehensive is 'creamed' by that grammar school. Given that a ward contains a fairly small number of secondary age pupils (typically 300 or so) and is usually geographically fairly small, we make the assumption that pupils in the same ward are essentially neighbours and face the same choices over which school to attend.

Notice that this does not assume that, in the absence of grammar schools, all pupils would simply attend the nearest comprehensive. The key supposition is that if there were no grammar schools, and their pupils were to be reallocated among all the local comprehensives, they would attend them in the same proportions as their neighbours in the same wards who currently attend comprehensive schools.

We therefore work through the process of taking all the pupils in grammar schools and reallocating them in this way to comprehensives. Each comprehensive can then see how many pupils it would gain under this process, or, to put it another way, how many it has 'lost' from the hypothetical total as a result of creaming by the grammar school.

In an area where there is more than one grammar school, we can aggregate to find the total number 'lost' to all grammar schools. We calculate this as a percentage of the total number who would have attended the school if there were no grammars; hence we have the extent to which that comprehensive is 'creamed' as a whole. A figure of, for example, '5% creamed' indicates that a school has lost to grammar schools 5% of the pupils it would have had. Here the phrase 'would have had' implies 'had there been no grammar schools and their current pupils had simply followed their neighbours to comprehensives'.

7.1.3. Example of the calculation of creaming

Consider a simplified unit in which there are two wards (*w1*, *w2*), two grammar schools (*g1*, *g2*) and three comprehensive schools (*c1*, *c2*, *c3*). The numbers of pupils in each combination of ward and school are shown in Table 32.

Table 32: Number of pupils in different wards and schools

Ward	School	Number of pupils
w1	*g1*	40
	c1	100
	c2	60
w2	*g1*	60
	g2	30
	c1	70
	c3	140

In ward *w1* there are 200 pupils, of whom 40 attend a grammar school, *g1*. Hence for this ward, 20% of pupils attend grammar schools. The 160 pupils attending comprehensives are divided between two schools, *c1* and *c2* in the ratio 5:3. If we reallocate the 40 grammar school pupils to the two comprehensive schools, *c1* and *c2*, they will get 25 and 15 respectively. Hence we would say, for example, that *c1* has 'lost' 25 pupils to *g1* from that ward.

In ward *w2* there are 300 pupils, of whom 90 (30%) attend grammar schools. If we reallocate them among the comprehensives, the 60 from *g1* are split 20:40 and the 30 from *g2* split 10:20 to *c1* and *c3* respectively. The numbers reallocated are shown in Table 33.

Table 33: Pupils reallocated from each grammar school

Ward	Comprehensive School	Number of pupils		
		already attending	reallocated from g1	reallocated from g2
w1	*c1*	100	25	0
	c2	60	15	0
w2	*c1*	70	20	10
	c3	140	40	20

If we look at comprehensive school *c1*, it has 170 pupils currently, but may be said to have 'lost' a further 45 to *g1* and 10 to *g2* as a result of creaming. Overall, this school has lost 55 out of a hypothetical pool of 225 pupils, so we would describe it as '24.4% creamed'.

7.1.4. The steps in the algorithm

The following steps describe in detail exactly how the calculation was done:

1. For the pupils attending each grammar school, identify the wards in which they live
2. For each such ward, calculate
    a. the number of pupils living there who attend that grammar school

   b. the total number of pupils living there who attend comprehensive schools;

3. Of these last (i.e. 2b), calculate the proportion who attend each comprehensive school.
4. Divide the number in 2a among the comprehensive schools in the proportions according to 3. These grammar school pupils from that ward are effectively reallocated to comprehensive schools in the same proportions as their neighbours in the same ward who attend those comprehensive schools.
5. For each comprehensive school and each ward from which it recruits we therefore have two numbers
   a. the number of pupils from that ward who attend the school
   b. the number of pupils from that ward who attend the grammar school, but would be 'reallocated' to that comprehensive school under the process in 4 (this may not be a whole number).
6. Add these two numbers together (5a + 5b) to calculate the 'pool' of pupils from that ward who would attend the comprehensive school if there were no grammar school.
7. Add up the numbers in 5b for all wards having pupils going to that comprehensive school. This is the total number of pupils 'lost' by that comprehensive to the grammar school.
8. Add up the total pool from all the wards with pupils going to that comprehensive school. This is the total number of pupils who would have attended that comprehensive school if there had been no grammar school and the pupils who currently attend the grammar school went to the different comprehensives in the same proportions as their neighbours.
9. Repeat steps 1 – 8 for all other grammar schools. Then for each comprehensive we have
   a. the total number lost to grammar schools as a whole
   b. the total pool that comprehensive would have had, if there had been no grammar schools
10. Calculate 9a as a percentage of 9b, to give the 'percentage creamed' for that comprehensive.

7.1.5. *Patterns of creaming in England*

Figure 8: Extent to which non-selective schools in England are creamed by grammar schools

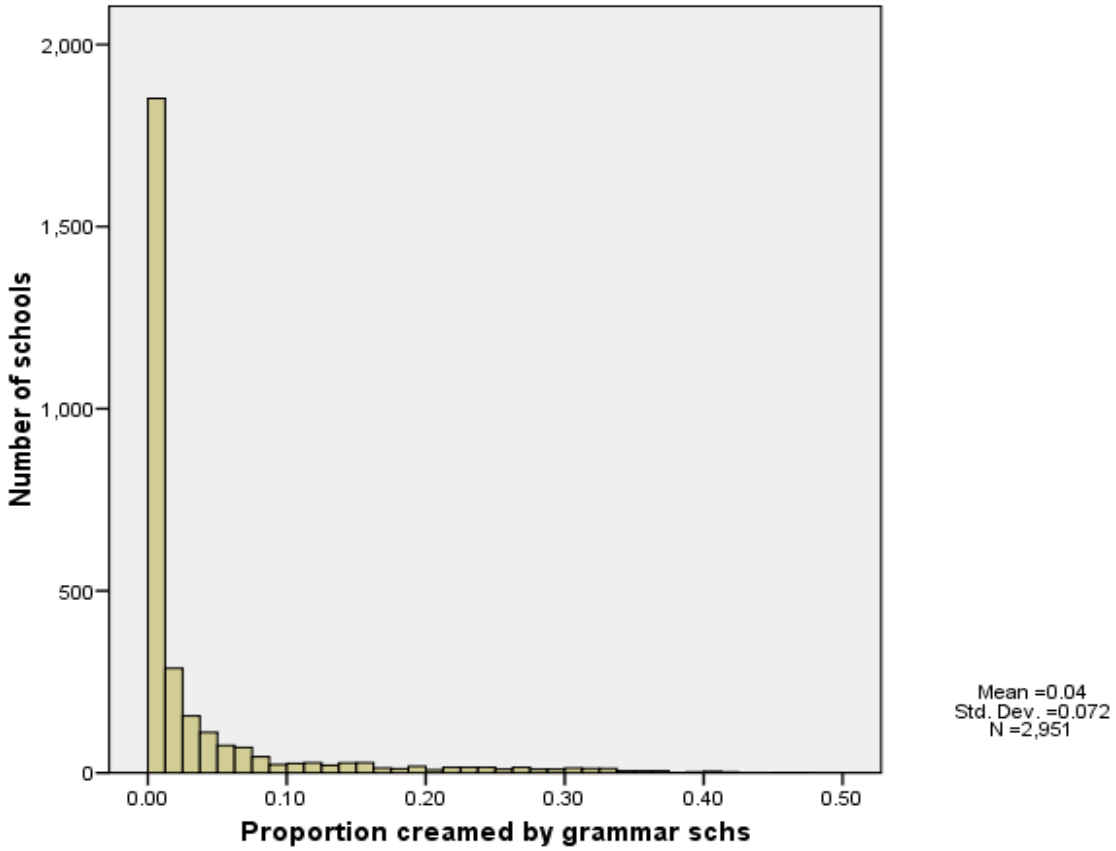


Figure 8 shows the distribution of the levels of creaming among non-grammar schools in England. It is apparent that very few schools seem to be heavily creamed, while the vast majority lose a very small number of pupils, if any, to grammar schools. Eight hundred and twenty six schools (28% of all non-grammar schools) lose no pupils at all to any grammar school. A further 32% lose fewer than 1% of pupils. About a third (35%) of non-grammar schools lose between 1% and 20%. Finally, only 161 schools (5%) lose more than 20% of their pupils to grammar schools. This last seems a surprisingly small number, given that it is fewer than the total number of grammar schools themselves, and that a typical secondary modern school would lose around 25% of its pupils. If the two types of school were the same size, we might expect to find three secondary moderns for every grammar school.

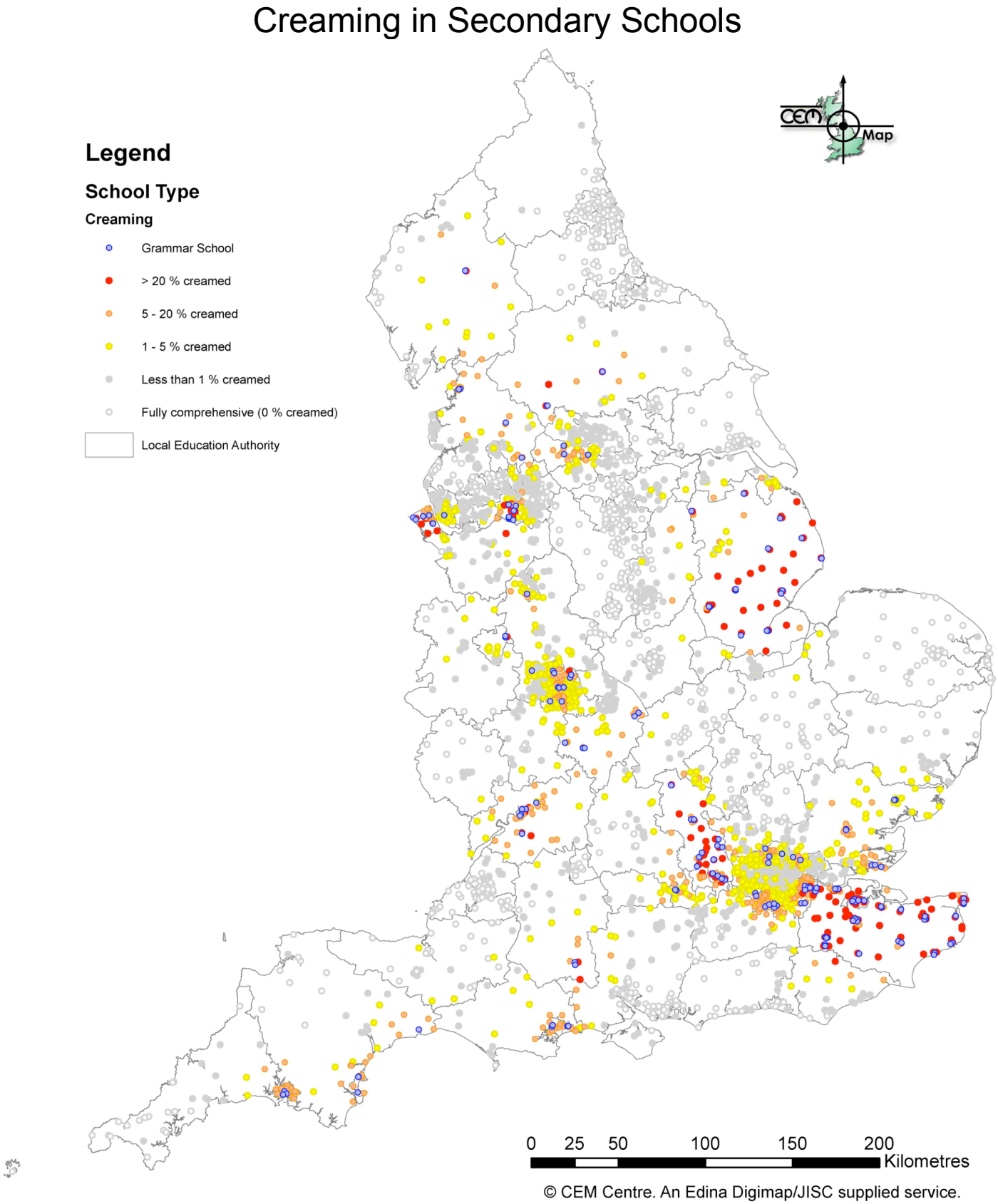
The areas in which these highly creamed schools are located are quite limited. A little over half of them (88 schools) are in either Kent or Lincolnshire. Two further LAs (Buckinghamshire and Medway) bring the total to 119. Hence almost three-quarters of all the schools that lose over 20% of their pupils to grammar schools are in just four LAs.

It is noticeable from Figure 8 that there is no clear cut-off that would enable us to distinguish between schools that are 'creamed' and those that are not. Our intention in examining creaming was that we would be able to identify schools that were affected by the existence of neighbouring selective schools, and compare them with those that were not. However, no two such

categories emerge. Although there are some schools that are completely unaffected by selection, these are a relatively small proportion (28%) of all non-selective schools in the country, given the small number of grammar schools. Some of those that are affected by creaming are affected so slightly that in practice the effect would be too small to notice. For example, a school that is less than 0.5% creamed would lose less than one pupil from a cohort of 200. This might be considered to have a negligible impact on its composition or overall performance, yet over a thousand schools in England fall in this category. As the rate of creaming rises above 0.5% it could start to be significant, but there is no distinct point at which such an impact would be evident.

Figure 9 shows the location of creamed schools across England. As would be expected, the areas of heavier creaming closely match the areas where there is a higher percentage of grammar school attendance (see Figure 3, p141) for example in Kent, Buckinghamshire, Lincolnshire, Gloucestershire, Wirral and areas of the West Midlands. Perhaps more surprising is the extent of low-level impact of grammar schools in some areas where there are not many of them. For example, in a ring around outer London there is hardly a school that does not lose 1% or more of its pupils to grammar schools.

Figure 9: Map of schools by percentage creamed (England)



7.1.6. *Do creamed schools really lose their cream?*

One suggestion from looking at the maps of creamed schools is that there seems to be a tendency for the schools that lose pupils to grammar schools to be themselves located in relatively affluent areas. In general, it looks as though schools in inner cities are losing fewer pupils than those in the suburbs. If this were the case, we might expect that even schools that were relatively heavily creamed might nevertheless maintain a reasonably balanced population; the loss of some of their most able pupils to grammar schools might be compensated by their capacity to recruit able pupils from their relatively socially advantaged catchments.

Figure 10: Proportions of able pupils in schools with different rates of creaming

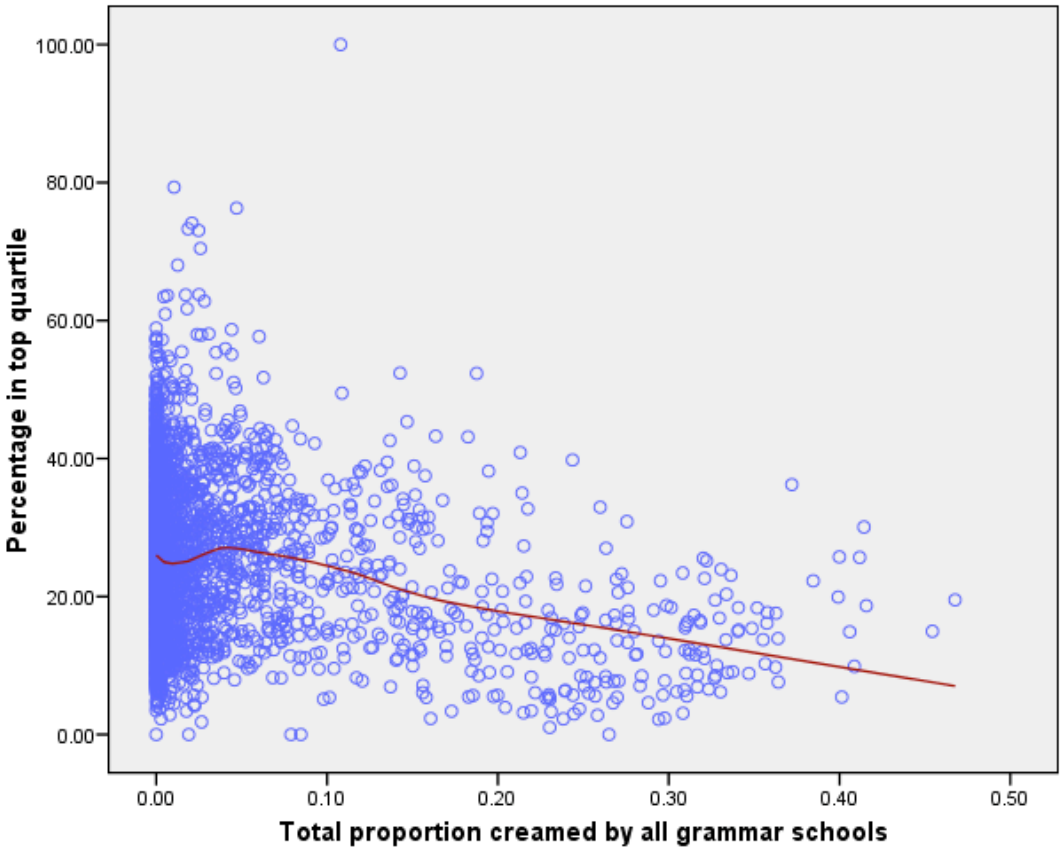


Figure 10 shows the relationship between the extent to which a non-selective school is creamed by grammar schools and the proportion of its pupils whose KS2 scores put them in the top quarter of ability nationally. The graph suggests that, for schools creamed up to about 10%, they do indeed maintain more or less their fair share of the most able pupils. Beyond 10% creaming, there is a slight fall off in the proportion of able pupils.

7.1.7. *Attempt to define selective systems of grammar schools*

We saw in Section 1.1.3 (p10) that LAs do not form distinct selective systems as there is considerable variation between the proportion of grammar schools, non-grammar schools and comprehensive schools in LAs with grammar schools. Section 6.3 (p142) also shows us that there is a considerable amount of boundary crossing between LAs by pupils attending grammar

schools, and hence that the LA is not an appropriate unit to analyse as a selective system. How, then, can we define selective systems?

If we remove the idea of LA boundaries and think about grammar schools and the schools that they cream from, are there particular groups of grammar and non-grammar schools that form isolated selective systems? In other words, are there groups of non-grammar schools that are creamed from one set of grammar schools and no others forming an enclosed system? We attempt to answer this question using factor analysis.

The factor analysis method works using the following reasoning. For any given grammar school we have calculated the creaming effect for each non-grammar. This means we can look at any pair of grammar schools, and see how far they cream from the same non-grammar schools. In fact, we can calculate a correlation between each pair of grammar schools, based on the level of correspondence of their creaming from each non-grammar school. High correlations mean that they are creaming from the same schools and they are therefore considered as being in the same cluster. Factor analysis is a technique that then puts the grammar schools into groups according to the correlations among them.

In an ideal situation, we would find that all the grammar schools in the same factor or group would be highly correlated with each other, and that any pair not in the same group would be uncorrelated. Unfortunately, we did not find this to be the case. In reality, we found that we could define groups to make the first condition true (that all in the same group were correlated) or we could define them to make the second true (that those in different groups would be uncorrelated), but no groupings could meet both conditions simultaneously. We have therefore failed to identify self-contained selective systems for grammar schools in England.

7.2. *Selectivity*

One of the arguments against grammar schools is that they are divisive, both socially and academically (see Chapter 2). However, it has also been argued that many 'comprehensive' schools are also highly socially segregated (Gorard *et al*, 2003). In this section we attempt to look at patterns of selection and segregation in both grammar and non-grammar schools.

In doing this, we consider the question raised in Section 6.1 (p137) as to whether the level of social selection evident in grammar school populations is higher than would be an inevitable consequence of their academic selection.

7.2.1. Defining selectivity

How 'selective' are non-selective schools? We would expect grammar schools to be selecting pupils based on academic ability, but might not expect non-grammar schools to be selective. If a non-selective school is over-subscribed it must apply admissions criteria of various kinds to decide which pupils will be offered places. The most commonly used criteria relate to where the pupils live, such as proximity of residence, catchment areas or feeder schools (Gorard *et al*, 2003, p143). All of these criteria are essentially location based; pupils in a given ward are equally likely to get a place at a given school. None of these should result in the more able pupils being selected from any given

ward. We also know that proximity is one of the main factors influencing pupils'/parents' choices about which school to apply to (Gorard, 1999)

If pupils' choice of preferred school and schools' selection processes are both essentially location based, then we would not expect there to be any differences between the pupils living in a given ward who go to one school and those who go to another. If they live in the same ward, we would expect them to be equally likely to apply to a particular school and, if they do apply, to get in. Hence, if we look at the pupils in a given school and the wards in which they live, we can compare the characteristics of these pupils with those in the same ward who go to other schools. Any differences in these characteristics we have defined here as *selectivity*.

We have created two scales for selectivity - one based on academic ability and the other based on social factors. The former uses the average of a pupil's marks in their three KS2 subjects; the latter, their FSM eligibility (coded as 1=eligible, 0=not eligible). In each case, the procedure for a particular school is as follows.

First split the pupils who attend that school into the wards in which they live. For each ward, calculate the average value (KS2 or FSM) for those pupils. Next, for each of these wards, identify any pupils who live in the same ward but attend another mainstream secondary school[6], and calculate their average. Subtract this from the previous average to get the difference between pupils who attend the school and those from the same ward who do not. These differences are then weighted by the number of pupils attending the school from that ward, and averaged across all wards. This average difference is then our index of selectivity.

In the case of FSM, there were two slight modifications to this. The first is that the direction was reversed so that a positive figure would mean that a school is taking in pupils who are on average more socially advantaged than the wards they come from. This gives our two indices of selectivity the same direction.

The second arose from the potential problems of comparing differences in proportions, particularly when many of the proportions eligible for FSM are close to zero. It could be argued, for example, that a school with 0% FSM is more different from the neighbouring rate of 5% than it would be if the corresponding figures were 20% and 25%. It would also mean that grammar schools in areas with high background FSM rates would be likely to appear more socially selective than those in areas of low FSM, since in the latter any differences between FSM rates are effectively capped.

We used a *logit* transformation to address this problem. This stretches out the scale at the ends to make the differences comparable. However, although this transformation did make the distribution of selectivity more Normal, it did not appear to improve the fit of any of our models, so we returned to the simple difference on the grounds of transparency.

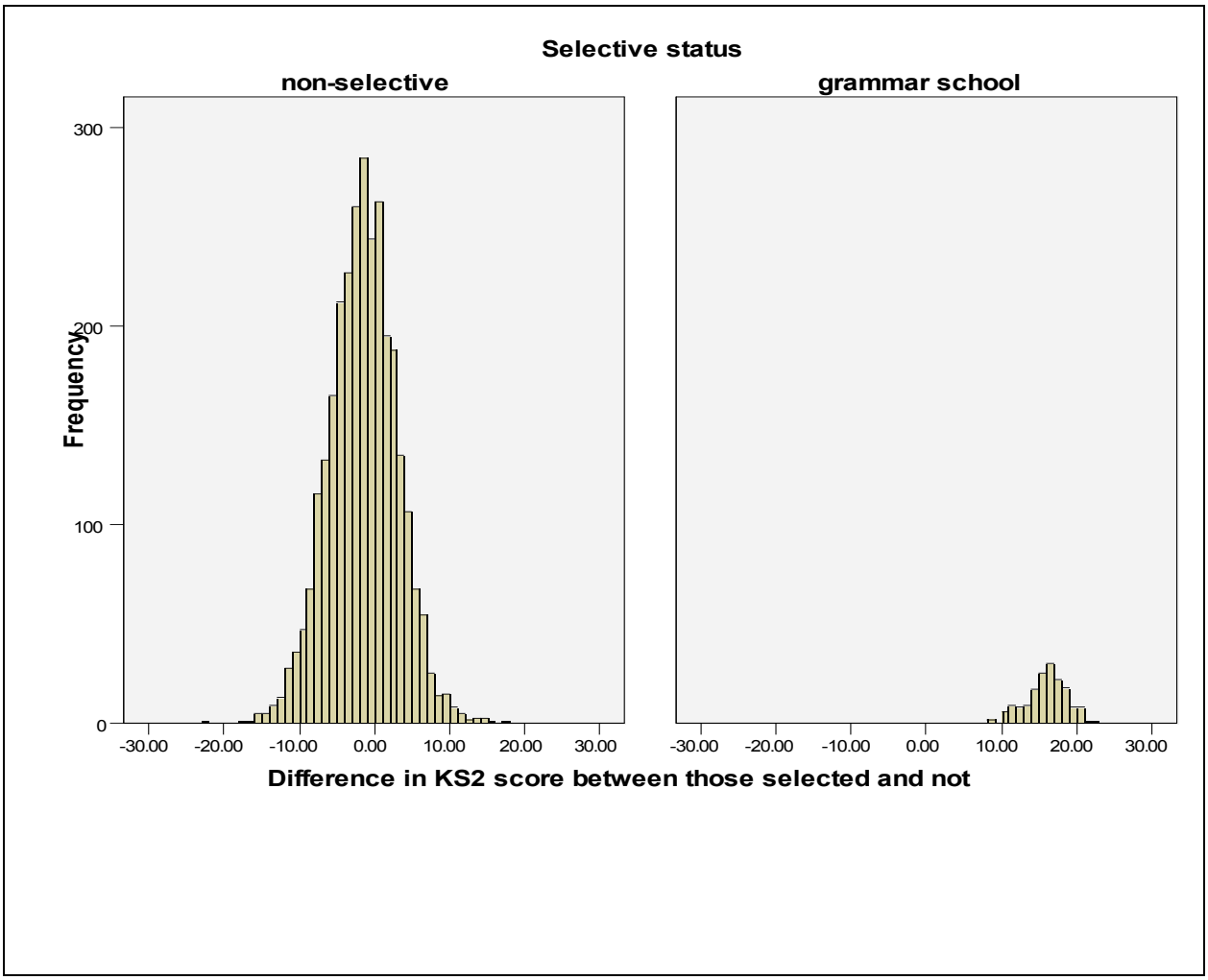
[6] In some cases there are no such pupils, and the pupils who attend the school from that ward are ignored in the calculation, since they have no neighbours with which to compare them.

7.2.2. *Selectivity in grammar and non-grammar schools*

The scales for selectivity defined in 7.2.1 have been used to produce Figure 11 and Figure 12, which compare the academic and social selectivity of grammar and non-grammar schools. The charts represent the data as histograms. In each chart the distributions for grammar and non-grammar schools are shown separately but using the same scale on the axes so they can be compared.

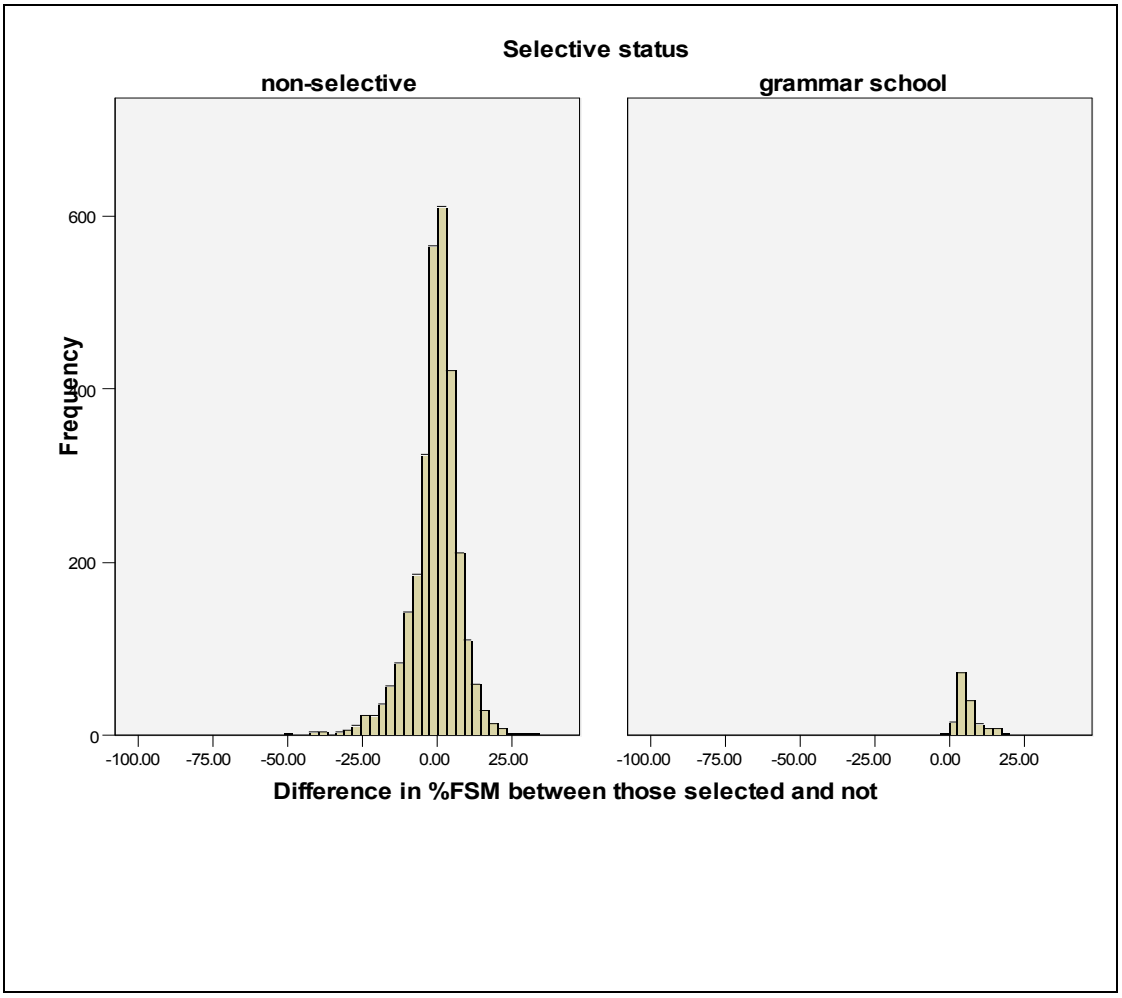
Figure 11 shows that nationally pupils selected by grammar schools have KS2 point scores typically 10-20 points higher than the average for the wards they are selected from. Given the selective nature of grammar schools, this would be expected. However, the charts also reveal that there are non-grammar schools that are being just as selective as grammar schools. Given that academic selection is the defining feature of grammar schools and the law expressly forbids other schools from selecting pupils in this way, this does seem rather surprising.

Figure 11: Comparison between academic selectivity of grammar and non-grammar schools in England



If all schools are ranked by their academic selectiveness, ten of the top 164 schools are non-grammar schools. The most selective non-grammar school is ranked 56 and has a difference of 17 marks between its pupils' average KS2 marks and those of their neighbours who attend other schools. The least selective grammar school is ranked 214th , with 50 non-grammar schools more academically selective than it. Interestingly, of these 50 non-grammar schools in the 'overlap' group, almost half (24) are faith schools, all with Voluntary Aided status. The remaining 26 non-faith schools comprise 3 Voluntary Aided, 7 Foundation schools, 8 Community schools, 6 CTCs and 2 Academies.

Figure 12: Comparison between social selectivity of grammar and non-grammar schools in England



Nationally the pattern of *social* selectivity is a little different (see Figure 12). Certainly, grammar schools are *on average* more socially selective than other

schools. That is to say, pupils who attend these schools are less likely to be eligible for free school meals than their neighbours who attend other schools. However, the charts show that the most socially selective schools in the country are not grammar schools but non-grammar schools who are somehow managing to attract or select the most socially advantaged pupils from the wards in which they recruit.

This too seems such a surprising result that it warrants some further investigation. If we rank all schools in order of their social selectivity, the top 100 most selective schools include 17 grammar schools. This is a significantly higher percentage than the proportion of grammar schools among all schools (5%), so grammar schools are certainly over-represented among the most socially selective schools. However, despite this, the vast majority (83) of the top 100 are not grammar schools. These proportions are shown graphically in Figure 13.

One reason why grammar schools are not at the top of this list is that they are typically situated in areas where the prevailing rates of FSM are below average; the average rate of FSM in wards from which pupils attend grammar schools is 8%, compared with 13% for non-grammars. The five most socially selective schools in the country all have FSM rates below 5%, despite rates of around 35% or higher in the catchments from which they draw. The highest ranked grammar school is number 24, with a 19 percentage point difference in FSM rates between its pupils and their neighbours in other schools.

Figure 13: Proportions of grammar and non-grammar schools among all schools, and among the most socially selective 100 schools

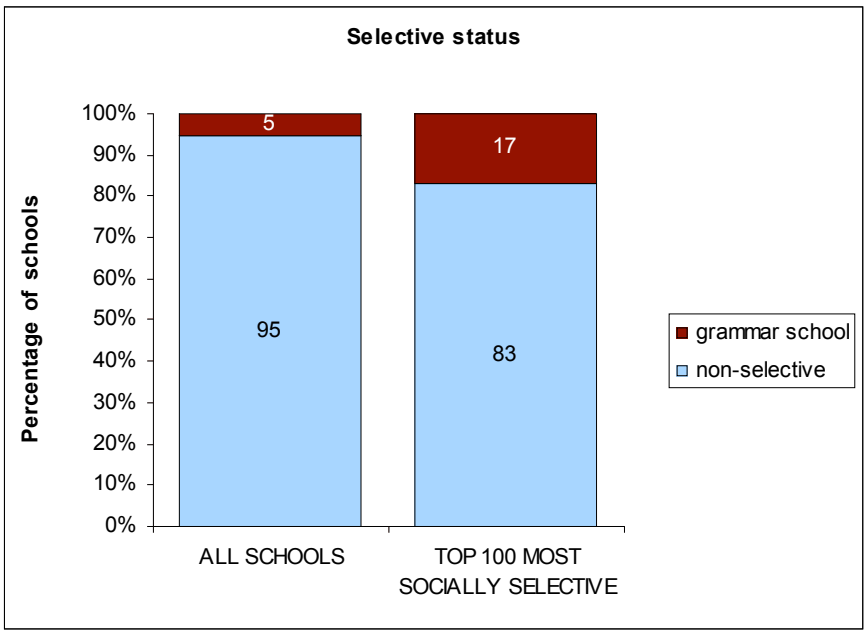


Table 34 graphically presents in more detail the characteristics of the top 100 most socially selective schools. It also makes a comparison with 100 randomly selected schools. We can see that the 100 most socially selective schools

include proportionally more grammar schools, more voluntary aided and less community schools, more faith schools and more single-sex schools than would be expected on average.

Table 34: Comparison of characteristics of the 100 most socially selective schools against 100 random schools

100 most socially selective schools								100 random schools						
Rank: Social selectivity	School % FSM	% FSM in catchment	% FSM difference	Selective status	Governance type	Faith status	School sex mix	School % FSM	% FSM in catchment	% FSM difference	Selective status	Governance type	Faith status	School sex mix
1	4	43	38	N	VC	CE	M	11	37	26	N	Co	No	M
2	0	36	36	N	Ot	No	M	42	36	-6	N	Co	No	B
3	5	37	32	N	VA	Ot	G	25	37	12	N	VA	RC	G
4	3	35	32	N	Co	No	M	31	48	17	N	VA	CE	M
5	3	34	31	N	VC	CE	M	30	24	-6	N	Fo	No	M
6	3	34	31	N	VA	CE	G	34	31	-3	N	VA	CE	M
7	10	38	28	N	Co	No	M	16	28	12	N	Co	No	M
8	11	37	26	N	Co	No	M	9	23	13	N	VA	RC	M
9	35	60	25	N	VA	RC	B	4	9	4	N	Fo	No	B
10	15	38	23	N	VA	CE	B	1	14	13	N	VA	CE	M
11	4	26	23	N	Co	No	M	3	4	1	N	Co	No	M
12	13	35	23	N	VA	CE	G	15	13	-2	N	Co	No	B
13	32	55	22	N	VA	RC	G	42	38	-4	N	Co	No	M
14	15	37	22	N	VA	RC	B	19	16	-3	N	Co	No	M
15	13	35	22	N	VA	Ot	M	25	20	-5	N	Fo	No	M
16	13	35	22	N	VA	RC	G	25	18	-7	N	Co	No	M
17	2	24	22	N	VA	CE	G	6	5	-1	N	Co	No	M
18	9	29	21	N	VA	RC	G	33	25	-8	N	Co	No	M
19	8	28	20	N	VA	RC	G	20	25	5	N	VA	RC	G
20	16	35	20	N	VA	RC	G	32	31	-1	N	VA	RC	G
21	13	33	20	N	VA	RC	B	37	27	-10	N	Co	No	M
22	9	29	19	N	VA	RC	M	42	31	-11	N	Co	No	M
23	17	36	19	N	VA	RC	B	48	40	-7	N	Co	No	M
24	8	27	19	N	Ac	No	M	48	21	-27	N	Co	No	M
25	3	22	19	G	VA	Ot	G	15	26	11	N	VA	RC	M
26	9	27	18	N	Co	No	M	24	24	1	N	Co	No	G
27	7	25	18	N	VA	CE	M	14	17	3	N	Co	No	M
28	5	23	18	N	VA	RC	B	16	11	-5	N	Co	No	M
29	6	24	18	N	Co	No	M	13	15	1	N	Co	No	M
30	16	34	18	N	Co	No	M	11	21	9	N	Co	No	M
31	3	22	18	G	VA	Ot	G	18	22	4	N	VA	RC	G
32	7	25	18	N	Fo	No	M	34	20	-14	N	Co	No	M
33	14	32	18	N	VA	RC	G	19	15	-4	N	Co	No	M
34	19	37	18	N	Co	No	B	22	20	-2	N	Co	No	M
35	2	19	17	G	VA	Ot	B	12	17	5	N	VA	RC	M
36	6	23	17	N	Fo	No	M	7	11	4	N	Co	No	M
37	4	21	17	N	VA	CE	B	18	21	3	N	Co	No	M
38	2	19	17	G	VA	Ot	M	5	7	2	N	Co	No	M
39	31	48	17	N	VA	CE	M	12	14	2	N	Co	No	M
40	1	18	16	G	VA	RC	G	2	2	0	N	Co	No	M
41	14	30	16	N	Co	No	M	3	2	-1	N	Co	No	M
42	19	35	16	N	Fo	No	B	6	11	5	N	VA	RC	M
43	5	21	16	N	VA	CE	M	2	1	-1	N	Co	No	M
44	10	26	16	N	Co	No	M	7	12	5	N	Co	No	M
45	24	40	16	N	Co	No	M	16	11	-5	N	Co	No	M
46	13	28	16	N	VA	RC	M	16	10	-6	N	Co	No	M
47	9	24	16	N	VA	RC	B	7	5	-3	N	Fo	No	M
48	18	33	16	N	VA	No	M	15	13	-3	N	Co	No	B
49	9	24	16	N	VA	RC	B	3	7	5	G	Fo	No	B
50	25	41	16	N	VA	RC	M	15	12	-3	N	Co	No	M
51	16	31	15	G	VA	No	B	11	19	8	N	Co	No	M
52	2	17	15	N	Fo	No	M	8	2	-6	N	Co	No	M
53	5	20	15	G	VA	No	M	8	9	1	N	Co	No	M
54	3	19	15	N	VA	CE	M	11	3	-8	N	Co	No	M
55	6	21	15	N	Co	No	M	6	4	-2	N	VC	CE	M
56	0	15	15	N	Co	No	M	17	19	2	N	Co	No	G
57	17	32	15	N	Fo	No	M	3	5	2	N	Co	No	M
58	9	24	15	N	Ac	No	M	5	1	-3	N	Co	No	M
59	1	16	15	N	Fo	No	G	7	7	0	N	Co	No	M
60	21	36	15	N	VA	No	B	3	2	-1	N	Fo	No	M
61	0	15	15	N	Co	No	M	1	5	3	N	VC	No	M
62	5	20	15	N	Co	No	M	10	11	1	N	Co	No	M
63	10	25	15	N	Fo	No	M	11	8	-4	N	VC	CE	M
64	0	15	15	N	VA	Ot	G	18	10	-8	N	Co	No	M
65	0	15	15	G	VA	No	G	14	3	-11	N	Fo	No	M
66	0	15	15	N	Co	No	M	9	4	-4	N	Co	No	M
67	8	23	15	G	VA	RC	B	2	10	8	G	Fo	No	B
68	2	16	15	G	VA	No	M	4	8	4	N	Fo	No	M
69	2	16	15	G	Co	No	B	6	10	4	N	Co	No	M
70	7	21	14	N	Fo	No	M	1	2	1	N	Co	No	M
71	4	18	14	N	VA	CE	M	3	5	2	G	Co	No	G
72	6	21	14	N	VA	CE	G	20	10	-10	N	Co	No	M
73	3	17	14	G	Fo	No	B	10	9	-1	N	VA	CE	M
74	1	15	14	G	VA	No	B	1	7	7	G	Fo	No	G
75	5	19	14	N	VA	CE	M	1	5	4	G	Co	No	B
76	3	18	14	N	Co	No	M	0	4	4	G	Fo	No	G
77	8	22	14	N	VA	RC	G	5	6	1	N	Co	No	M
78	11	26	14	N	Co	No	M	6	5	-1	N	Co	No	M
79	9	22	14	N	VA	RC	M	5	1	-3	N	Co	No	M
80	21	35	14	N	Co	No	M	10	9	-2	N	Co	No	M
81	3	17	14	N	VA	RC	M	8	0	-8	N	Co	No	M
82	22	36	14	N	VA	CE	M	0	3	3	G	VA	No	M
83	16	29	14	N	Fo	No	M	1	13	12	N	VA	RC	M
84	7	22	14	N	Fo	No	M	9	5	-4	N	Co	No	M
85	1	15	14	G	Fo	No	M	5	5	0	N	Co	No	M
86	6	19	13	N	Co	No	M	4	4	0	N	Co	No	M
87	9	23	13	N	VA	RC	M	4	6	1	N	Fo	No	M
88	1	14	13	N	VA	CE	M	1	6	5	G	Fo	No	M
89	11	24	13	N	Co	No	M	3	4	1	N	Fo	No	M
90	10	23	13	G	VA	RC	G	8	1	-7	N	Co	No	M
91	4	17	13	G	Fo	No	B	8	7	0	N	Co	No	M
92	0	13	13	N	VA	Ot	B	3	10	7	N	Co	No	B
93	19	32	13	N	VA	RC	G	5	3	-1	N	Co	No	M
94	22	35	13	N	VA	Ot	B	6	10	5	N	Co	No	M
95	4	17	13	N	Co	No	M	34	8	-25	N	Co	No	M
96	8	21	13	N	Fo	No	M	7	6	-1	N	Co	No	M
97	7	20	13	N	VA	RC	M	7	7	-1	N	Co	No	M
98	1	14	13	N	VA	CE	M	6	2	-4	N	Co	No	M
99	2	15	13	N	VA	RC	M	3	6	3	N	VC	CE	G
100	4	17	13	N	VA	RC	G	2	7	5	N	VA	CE	M

key

Selective status
- N non-grammar
- G grammar

Governance type
- VA voluntary aided
- VC voluntary controlled
- Co community
- Fo foundation
- Ac academy
- Ot other

Faith status
- No no affiliation
- RC Roman Catholic
- CE Church of England
- Ot other

School sex mix
- M mixed
- G girls only
- B boys only

7.2.3. *Accounting for selectivity in non-selective schools*

These results show that non-selective schools also appear to be 'selective'. That is to say, they are somehow attracting or selecting the 'best' pupils (in terms of ability or socioeconomic status) from a given ward. However, we cannot tell how the selectivity is taking place. One hypothesis is that oversubscribed schools are using an admissions policy that is based on criteria other than distance to school. Another hypothesis is that parents of different types of pupils living in the same areas (e.g. the more able), or indeed the pupils themselves, are choosing to apply to different types of school. For example parents of more able pupils may want to apply to schools that have more able pupils, or parents who have access to transport may be making different choices from those who do not. The following section attempts to investigate these ideas further, though ultimately this question goes beyond the scope of this report.

7.2.4. *Characteristics of the 100 most socially selective non-grammar schools*

It seems hard to understand how a non-selective school can have a 30 percentage point difference between the FSM rate in its catchment and the rate for the pupils it takes. What kinds of schools are these highly socially selective non-grammar schools? To answer this, we again ranked all non-grammar schools in order of their social selectivity, and compared the characteristics of the top 100 in this list with the national picture. As well as the selective status of a school (grammar school or not), other factors such as the school's governance type, faith status and sex mix seem to be obvious ones to consider in relation to social selectivity. Figure 14 to Figure 16 show this comparison, in terms of the governance type, faith status and sex mix of the school.

Figure 14: Proportions of different governance types among all non-grammar schools, and among the 100 most socially selective non-grammar schools

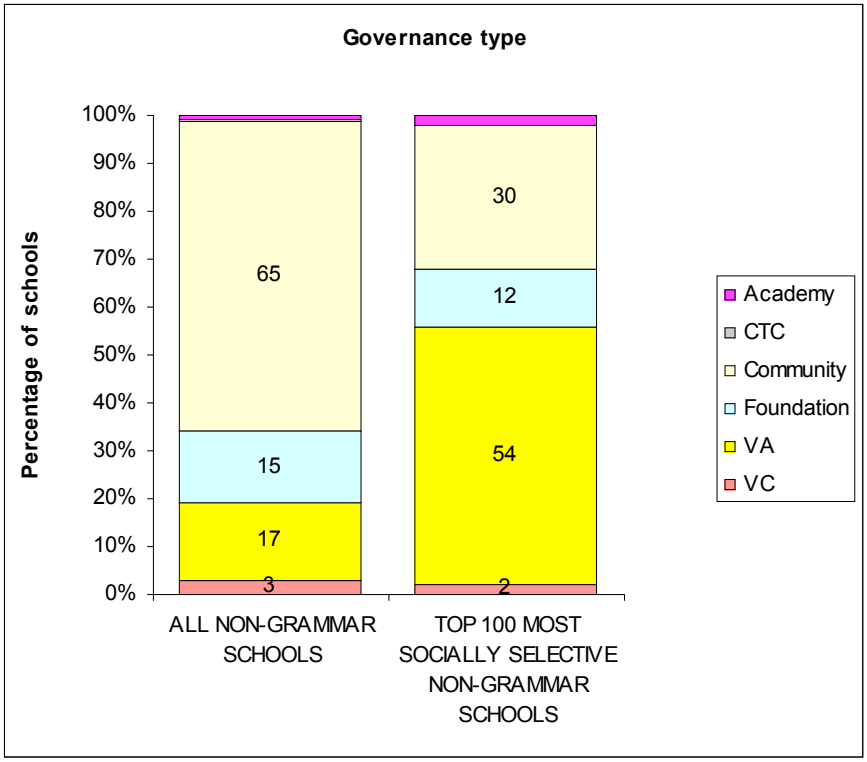
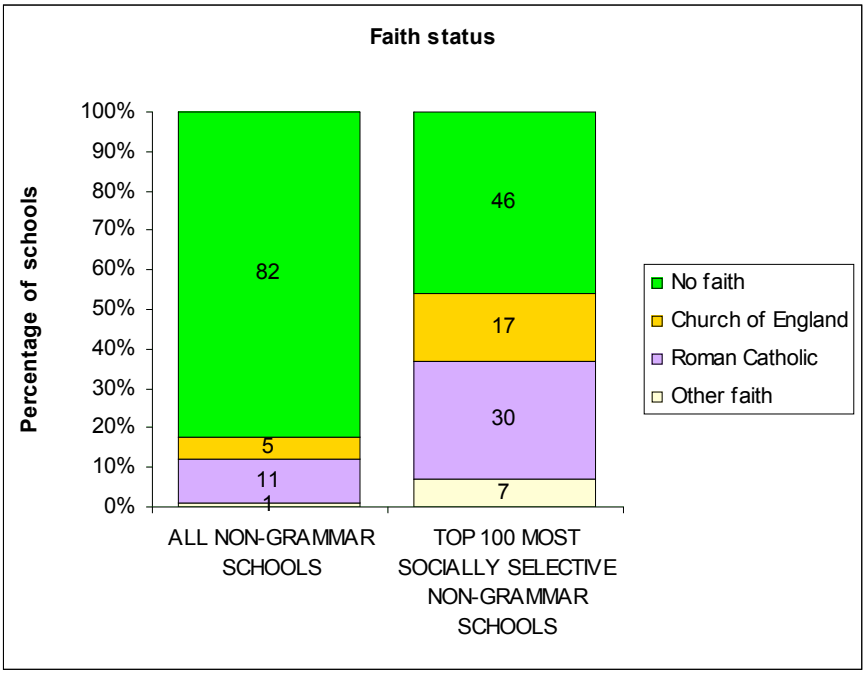


Figure 14 shows some clear differences in the governance of the most socially selective schools. The most striking is the dramatic over-representation of voluntary aided schools: 54 of the 100 most socially selective are VA, compared with 17% in the national population. Community schools, which form the majority of non-grammar schools are similarly under-represented in the socially selective group.

Figure 15: Proportions of faith types among all non-grammar schools, and among the 100 most socially selective non-grammar schools



In Figure 15 we see the proportions of different faith schools. In England as a whole, only 18% of non-grammar schools have any faith affiliation, but 54% of the most socially selective schools are faith schools. Church of England, Roman Catholic and other faith schools are all significantly over-represented in the 100 most socially selective schools.

Figure 16: Proportions of different sex mixes among all non-grammar schools, and among the 100 most socially selective non-grammar schools

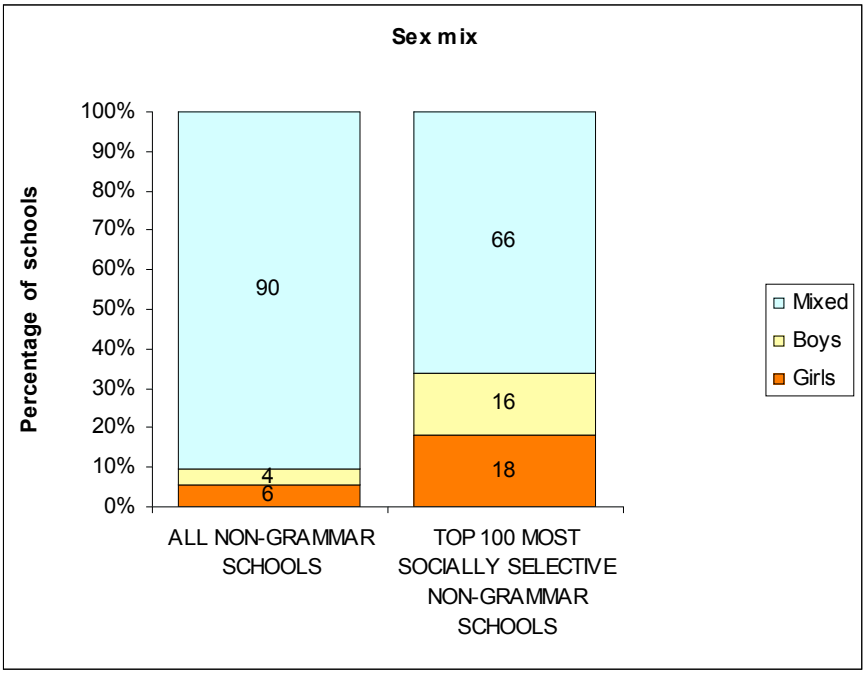


Figure 16 shows the split for single- and mixed-sex schools. Both boys-only and girls-only schools are over-represented among the most socially selective group. Single-sex schools account for 34 of the 100 most selective, but only 10% of the national population of non-grammar schools.

7.2.5. Average selectivity of different types of school

An analysis of the mean value of our two selectivity indices for schools in each of these categories (school's governance type, faith status and sex mix) is shown in Table 35. The table also shows the number of schools in each subgroup, together with standard deviations and standard errors.

For academic selectivity, the selective status of the school is, not surprisingly, the factor that makes the most difference, with over 17 KS2 marks difference between grammar schools and others. Selectivity also varies by school governance, with City Technology Colleges the most selective, followed by Voluntary Aided, Foundation and Voluntary Controlled schools. Least selective are Community schools and Academies. Although some of these groups have quite small numbers, the differences are enough to be statistically significant.

The faith status of the school appears to be related to its academic selectivity, with schools of faiths other than Church of England or Roman Catholic[7] the most selective, followed by Roman Catholic schools, Church of England schools, with those of no faith the least selective. The final grouping concerns the sex mix of the school, with boys only and girls only schools both around 6 points more selective than mixed sex schools.

[7] The majority of these are other Christian denominations, with a small number of Jewish (8), Muslim (2) and Sikh (1) schools.

Table 35: Social and academic selectivity statistics for different types of schools

		Difference in avg KS2 between those selected and not				Difference in %FSM between those selected and not			
		Mean	N	Std. Devn.	Std. Err.	Mean	N	Std. Devn.	Std. Err.
Selective status	non-selective	-1.52	2933	4.54	0.08	-0.87	2936	8.29	0.15
	grammar school	15.95	164	2.67	0.21	6.25	164	3.88	0.30
Type	Academy	-3.76	24	7.13	1.46	-12.48	24	15.43	3.15
	CTC	7.72	11	5.09	1.53	5.25	11	4.70	1.42
	Community	-2.08	1932	4.62	0.11	-2.22	1934	8.10	0.18
	Foundation	1.04	521	8.07	0.35	1.08	521	6.55	0.29
	VA	2.96	518	5.33	0.23	4.53	519	7.47	0.33
	VC	0.98	91	6.92	0.73	1.28	91	6.67	0.70
Faith	No faith	-1.17	2564	5.92	0.12	-1.41	2566	8.04	0.16
	Church of England	0.87	162	6.01	0.47	2.36	163	9.62	0.75
	Roman Catholic	2.45	334	3.70	0.20	4.47	334	6.48	0.35
	Other faith	5.32	37	9.28	1.52	6.37	37	9.83	1.62
Sex mix	Mixed	-1.39	2690	5.04	0.10	-1.08	2693	8.10	0.16
	Boys	4.71	182	8.77	0.65	3.67	182	7.86	0.58
	Girls	4.58	225	7.89	0.53	3.22	225	8.71	0.58

An obvious problem with the kind of analysis presented in Table 35 is that some of these categories intersect. As noted above, a disproportionate number of grammar schools are single sex, for example. Does the apparent selectiveness of single sex schools arise purely from the fact that many of them are in fact grammar schools, or is there an additional effect of being a single sex school after taking account of their selective status? To answer this kind of question we need to consider the different combinations of the intersections of these factors.

The numbers of schools in each of the different possible permutations of selective status, sex mix, faith status and governance type are shown in Table 36. The mean values of our indices of academic and social selectivity for each of these same subgroups are shown in Table 37 and Table 38, respectively.

Table 36: Number of schools in each combination of school selective status, sex mix, faith status and governance type

Selective status	Sex mix	School faith	School type					
			Academy	CTC	Community	Foundation	Voluntary Aided	Voluntary Controlled
non-selective	Mixed	No faith	15	11	1761	386	16	40
		Church of England	2			7	100	32
		Roman Catholic	1				259	
		Other faith	5				14	3
	Boys	No faith	1		51	24	5	3
		Church of England					7	
		Roman Catholic				1	28	
		Other faith					2	
	Girls	No faith			83	23	4	
		Church of England					10	2
		Roman Catholic					38	
		Other faith					3	
grammar school	Mixed	No faith			9	23	3	1
		Roman Catholic					2	
		Other faith					2	2
	Boys	No faith			7	27	10	6
		Church of England					4	
		Roman Catholic					2	
		Other faith				1	3	
	Girls	No faith			23	29	2	3
		Roman Catholic					3	
		Other faith					2	

Table 37: Mean value of academic selectivity for schools in each combination of school selective status, sex mix, faith status and governance type

Selective status	Sex mix	School faith	School type					
			Academy	CTC	Community	Foundation	Voluntary Aided	Voluntary Controlled
non-selective	Mixed	No faith	-2.8	7.7	-2.5	-1.9	2.2	-0.2
		Church of England				-0.9	0.8	-2.7
		Roman Catholic					2.1	
		Other faith	-6.0				2.8	-0.3
	Boys	No faith			-2.9	-1.4	3.0	-1.5
		Church of England					4.6	
		Roman Catholic					2.2	
		Other faith						
	Girls	No faith			-1.4	-0.3		
		Church of England					6.2	-0.9
		Roman Catholic					3.3	
		Other faith						
grammar school	Mixed	No faith			15.2	18		
		Roman Catholic						
		Other faith						
	Boys	No faith			14.5	15.4	15.4	15.3
		Church of England						
		Roman Catholic						
		Other faith						
	Girls	No faith			15.1	15.9		
		Roman Catholic						
		Other faith						

NB: values suppressed if fewer than 5 schools in each cell.

Table 38: Mean value of social selectivity for schools in each combination of school selective status, sex mix, faith status and governance type

Selective status	Sex mix	School faith	School type					
			Academy	CTC	Community	Foundation	Voluntary Aided	Voluntary Controlled
non-selective	Mixed	No faith	-13.2	5.2	-2.4	-0.1	3.7	1.2
		Church of England				2.9	2.1	0.4
		Roman Catholic					3.3	
		Other faith	-10.6				7.7	
	Boys	No faith			-1.6	1.2	7.7	
		Church of England					8.2	
		Roman Catholic					8.5	
		Other faith						
	Girls	No faith			-2.0	2.5		
		Church of England					7.6	
		Roman Catholic					8.2	
		Other faith						
grammar school	Mixed	No faith			4.3	6.8		
		Roman Catholic						
		Other faith						
	Boys	No faith			5.6	6.1	5.9	4.2
		Church of England						
		Roman Catholic						
		Other faith						
	Girls	No faith			5.4	5.3		
		Roman Catholic						
		Other faith						

NB: values suppressed if fewer than 5 schools in each cell.

The picture presented by Table 37 and Table 38 is complex and hard to make sense of. One way we can try to isolate the effects of the different factors is to use multiple regression. The next section presents this analysis.

7.2.6. *Modelling the relationships between selectivity and school characteristics*

A multiple regression model allows us to investigate the effects of all the different factors simultaneously. By coding each school with a set of 'dummy' variables to represent each factor, we can enter them into a regression model

that estimates the effect of each on our outcome measure (academic or social selectivity), whilst holding the effects of all the other variables constant.

A number of combinations of factors were investigated in order to achieve the best fitting model. These included some additional school characteristics related to the size of the school and its catchment. The first of these characteristics was simply the size of the school cohort. With all the school type factors in the model, cohort size was found to account for additional variance and had a statistically significant relationship with both academic and social selectivity. In both cases larger schools were more selective. In order to be able to compare the size of the effect, cohort size was dichotomised into a binary variable[8]; schools with fewer than 183 pupils (the median number) in the cohort were characterised as 'small', those with 183 or more, 'large'. The regression coefficients then estimate the difference between the selectivity of small and large schools, after taking other factors into account.

The second additional characteristic was the number of wards from which a school draws its pupils. This was intended to be an index of the size of the catchment area of a school. If schools are drawing their pupils from further afield than their local wards then they may be selecting on reasons other than distance. In the absence of a direct measure of distance, we considered two ways of judging from how far afield a school could be drawing its pupils:

i) The number of wards a school draws from
ii) The number of wards a school draws from, excluding those wards from which only one pupil comes

Whilst the first option is the most obvious choice, a school may have a single pupil from a ward due to any number of reasons (e.g. moving house), the second option counts the wards from which the school draws the bulk of its pupils.

We considered the possibility that wards in leafy suburbs might be less densely populated so a school may be drawing from a larger number of wards for no other reason than to fill the school. In order to discount this we looked at the correlation between the number of wards each school draws from and the average number of pupils in those wards for each school. The correlation is very weak (–0.079) which indicates that ward size appears to be unrelated to the number of wards a school is drawing from and therefore there must be other reasons why they are drawing from more wards. In particular this suggests that the number of wards a school draws from is a fair measure of the school's catchment area.

For the purposes of the regression model, the number of wards from which a school takes more than one pupil was again dichotomised to create a binary 'dummy' variable. The median number of wards was 12, so schools drawing at least two pupils from more than 12 wards were categorised as drawing from a large number of wards.

The third and final additional factor was the extent to which a school draws pupils from wards in which that school may be regarded as a minority destination. This was intended to be an index of the amount of competition

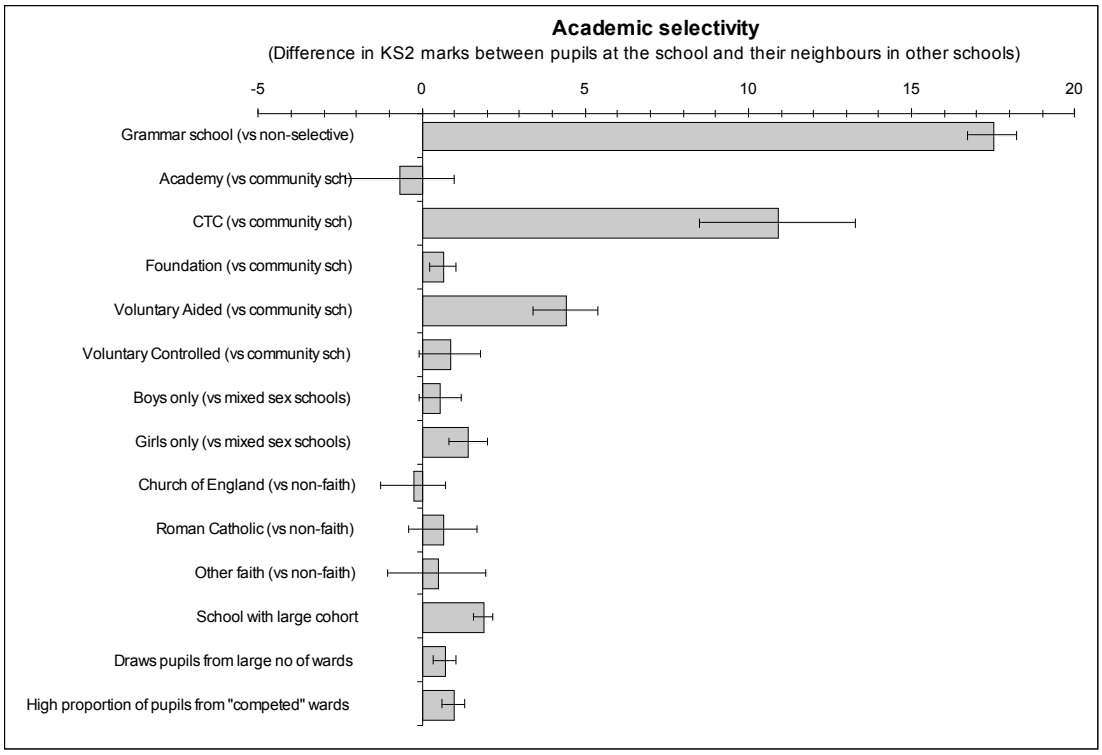
[8] In general, turning a continuous variable such as cohort size into a dichotomy (small/large) throws away information and reduces the fit of the model. However, as all the other variables in the model were dichotomies, in this case it also makes it easier to compare the effect of size with effects of other characteristics. A small reduction in model fit was judged to be acceptable in return for this benefit.

among schools. On average, secondary schools in England draw pupils in a given cohort from 23 wards. However, nine of those wards contribute only one pupil to that school; only eight wards contain at least five pupils. In most schools, the majority of pupils come from a relatively small number of wards; half come from wards from which at least 28 pupils attend that school. The average ward contains 73 pupils in a cohort, who attend nine different schools.

The index used was the percentage of pupils in a school who come from wards in which fewer than 20% of the pupils attend that school. Clearly, this overlaps somewhat with the number of wards from which the schools draws pupils: both are measuring an aspect of catchment size and competition (the correlation between the two is 0.57). Nevertheless, both factors explain unique variance in the model, so both were included. One again, this index was dichotomised at the median value, 69%.

Figure 17 and Figure 18 show the regression coefficients for the models for academic and social selectivity, respectively.

Figure 17: Estimates of the effects of different factors on academic selectivity, from multiple regression



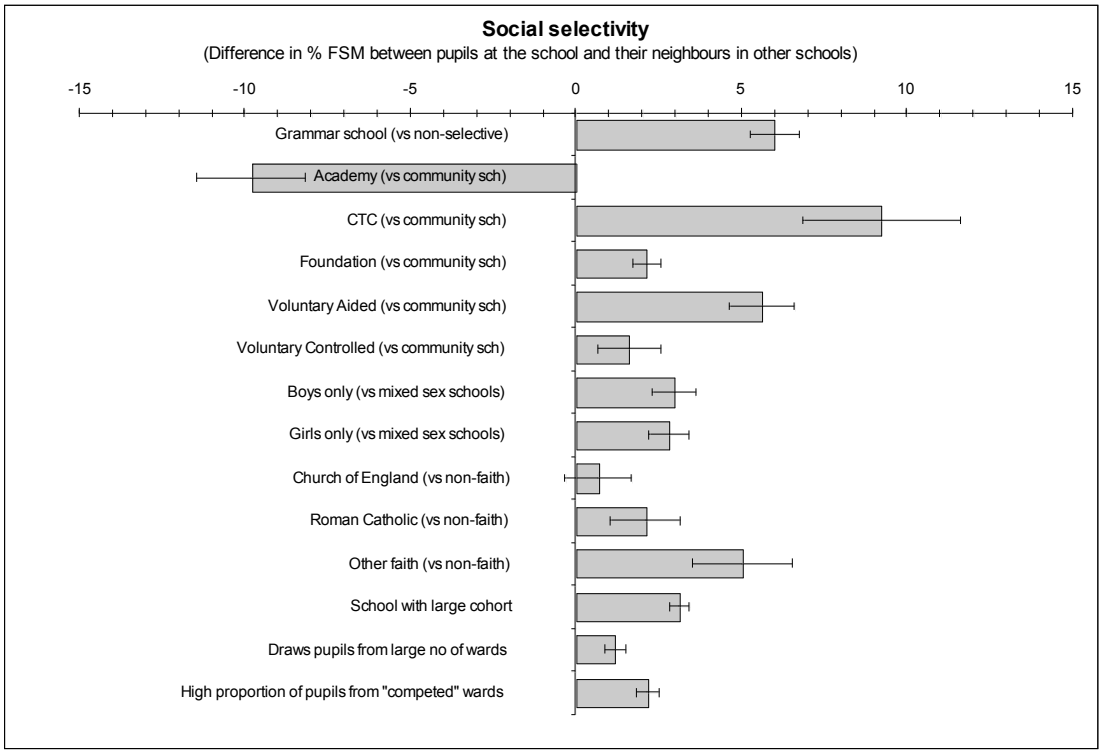
For academic selectivity (Figure 17) the model explains a good proportion of the variation ($R^2 = 0.55$). Not surprisingly, the largest difference in selectivity is associated with being a grammar school; this difference is unchanged by the addition of all the other factors in the model. What is surprising is the size of the effect associated with being a City Technology College (CTC), equivalent to an average difference of 11 KS2 marks between the pupils attending these schools and those in the same neighbourhoods attending other schools, by comparison with community schools. Although there are

only 11 CTCs with pupils in the 2006 GCSE cohort this difference is big enough and consistent enough for it to be highly statistically significant.

The next biggest difference is for voluntary aided (VA) status. This factor accounts for a difference of four KS2 marks in selectivity, compared to community schools. Although the vast majority (92%) of VA schools are faith schools and they are also more likely than other types to be single-sex and grammar schools, the 'voluntary aided' effect remains even after taking account of these characteristics.

Other factors associated with greater academic selectivity are being a large school or a girls-only school. Remaining factors have effects that are either smaller or not statistically significant. In particular, the apparent effects of faith status and being single-sex (especially boys only) seen in Table 37 have been substantially reduced by the inclusion of the other factors in the model.

Figure 18: Estimates of the effects of different factors on social selectivity, from multiple regression



The factors associated with the social selectivity of a school are shown in Figure 18. Here the model fit is rather less good than for academic selectivity, with an R^2 value of 0.20. Being a grammar school is associated with a six point difference, which has been reduced slightly by the inclusion of the other factors. The 'CTC' effect has increased in this model and accounts for nine percentage points greater selectivity than the baseline group, community schools. Academies are still looking significantly less socially selective than all other types for the 2006 cohort. Voluntary aided status accounts for almost a six point difference in selectivity. Both voluntary controlled and foundation schools are a little more selective than community schools; the effect of both factors has been reduced a little by the inclusion of the full model.

Being a single-sex school, of either sex, is still associated with being more socially selective than mixed schools, though again the difference has been reduced by the other factors. The association between faith status and selectivity has been reduced more than any other group of factors in this model. Being a Roman Catholic school and, even more so, a school of 'other faith' is associated with social selectivity, but the 'Church of England' effect is not statistically different from zero.

Additional factors in the model – school size, size of catchment (number of wards) and extent of competition –are all associated with selectivity and have helped to explain some of the variation that was previously associated with the various subgroups. Large schools are about three percentage points more socially selective than small; schools whose pupils come from competed wards are about two points more selective than those where pupils come from areas in which that school is the main destination; schools that draw from many wards are about one point more selective than those with smaller catchments.

7.2.7. Relationship between social and academic selectivity

Are grammar schools more socially selective than we would expect?

One way to answer this question is to compare the social and academic selectivity of both grammar and non-grammar schools. Such a comparison is shown in Figure 19.

Figure 19 shows that there is a moderate relationship between the two kinds of selectivity; schools that are academically selective tend also to be somewhat socially selective. In fact, the correlation between social and academic selectivity for non-grammar schools is 0.52. For grammar schools the figure is 0.27,[9] indicating that the relationship is considerably weaker, though this may simply reflect the narrower variation within that group.

It is clear from Figure 19, however, that, for a given level of academic selectivity, grammar schools are appreciably less socially selective than non-grammar schools. Hence, although grammar schools tend to be fairly socially selective, in the sense that the pupils who attend them have lower rates of FSM than others in the same neighbourhoods, they appear to be less socially selective than non-grammar schools with comparable levels of academic selection.

[9] The corresponding figures for the correlations using the logit transformed social selectivity index are 0.51 and 0.23 respectively.

Figure 19: Scatter graph of academic vs. social selectivity for grammar and non-grammar schools

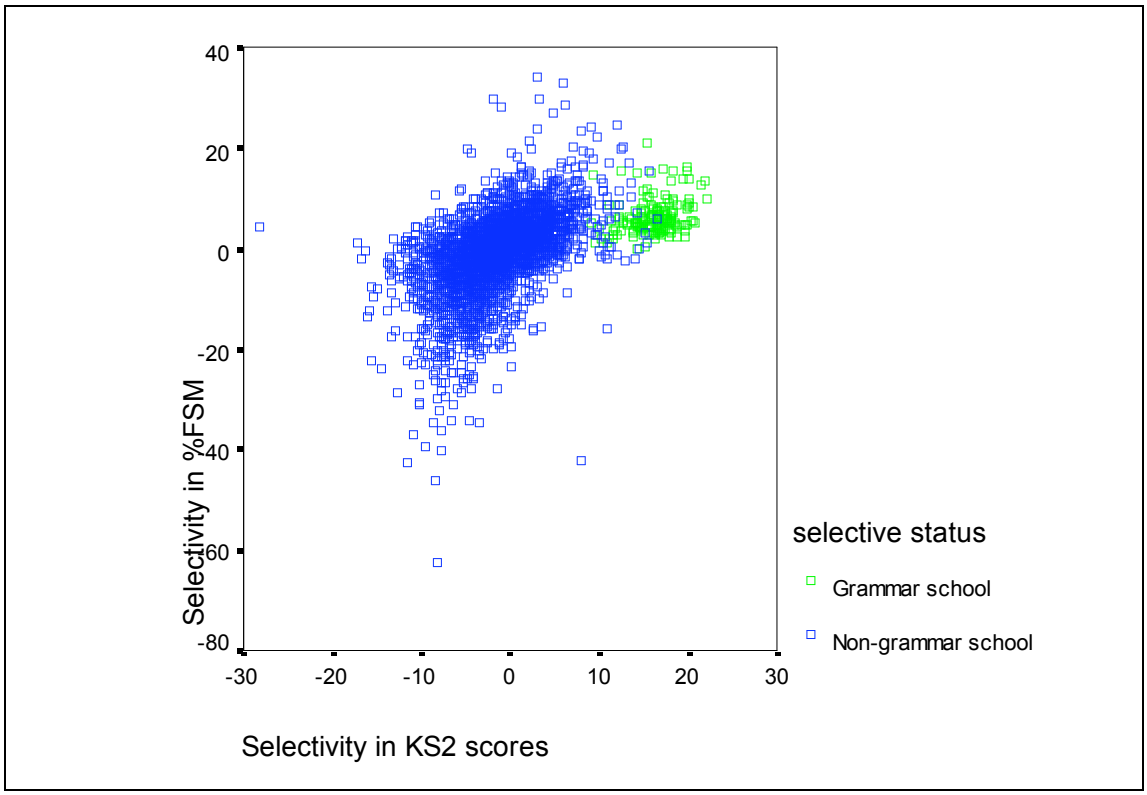
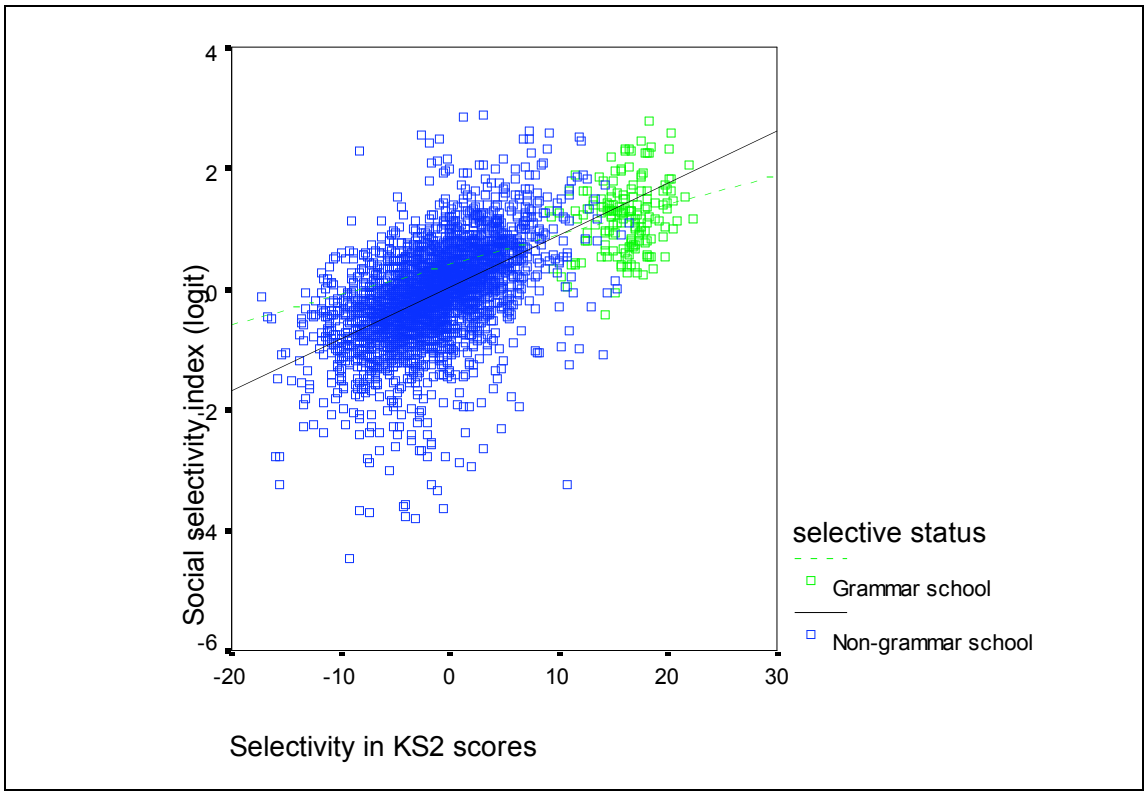


Figure 20 shows the same graph, but with Free School Meals selectivity transformed into logits. The regression line for each group is also shown. We can see that grammar school levels of social selectivity are now more similar to those of non-grammar schools with the same levels of academic selectivity. This is probably because the logit scale provides a better measure of social selectivity for schools in low FSM neighbourhoods than the capped measure derived from the raw difference. Nevertheless, even this scale shows the grammar schools as less socially selective than might be expected from their level of academic selection

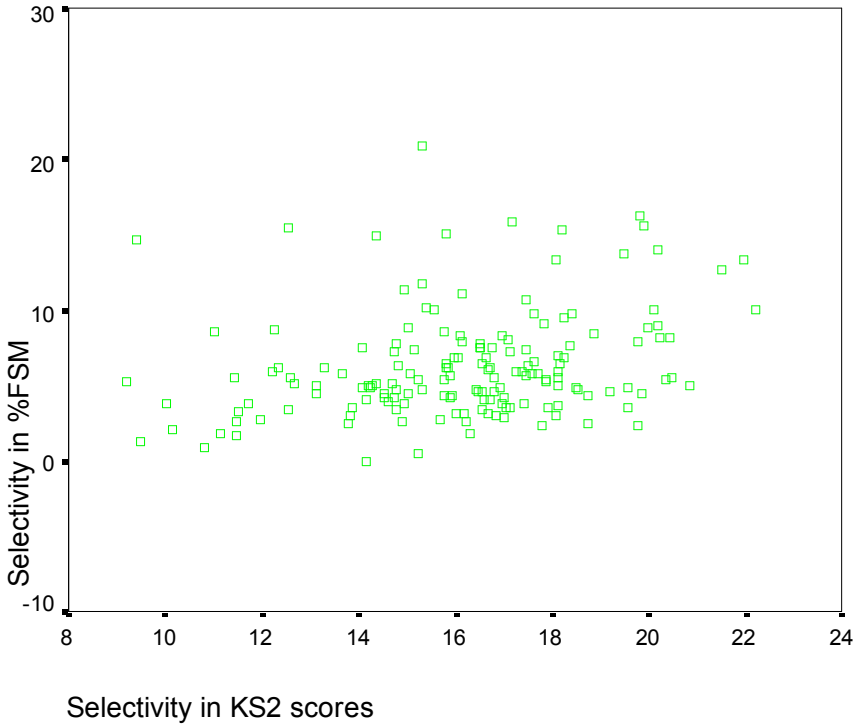
Figure 20: Scatter graph of academic vs. logit-transformed social selectivity for grammar and non-grammar schools (with regression lines shown)



7.2.8. *Social selectivity of individual grammar schools*

Figure 21 shows the relationship between social and academic selectivity for the grammar schools only. It is clear that grammar schools diverge considerably on the former variable; some take about the same proportion of FSM pupils as other schools around them, while others take a substantially smaller proportion. Yet, the weakness of the correlation suggests that the variation in different grammar schools' levels of social selectivity cannot be explained by how academically selective they are. Hence it is appropriate to ask what other factors might explain this variation.

Figure 21: Scatter graph of academic vs social selectivity (grammar schools only)



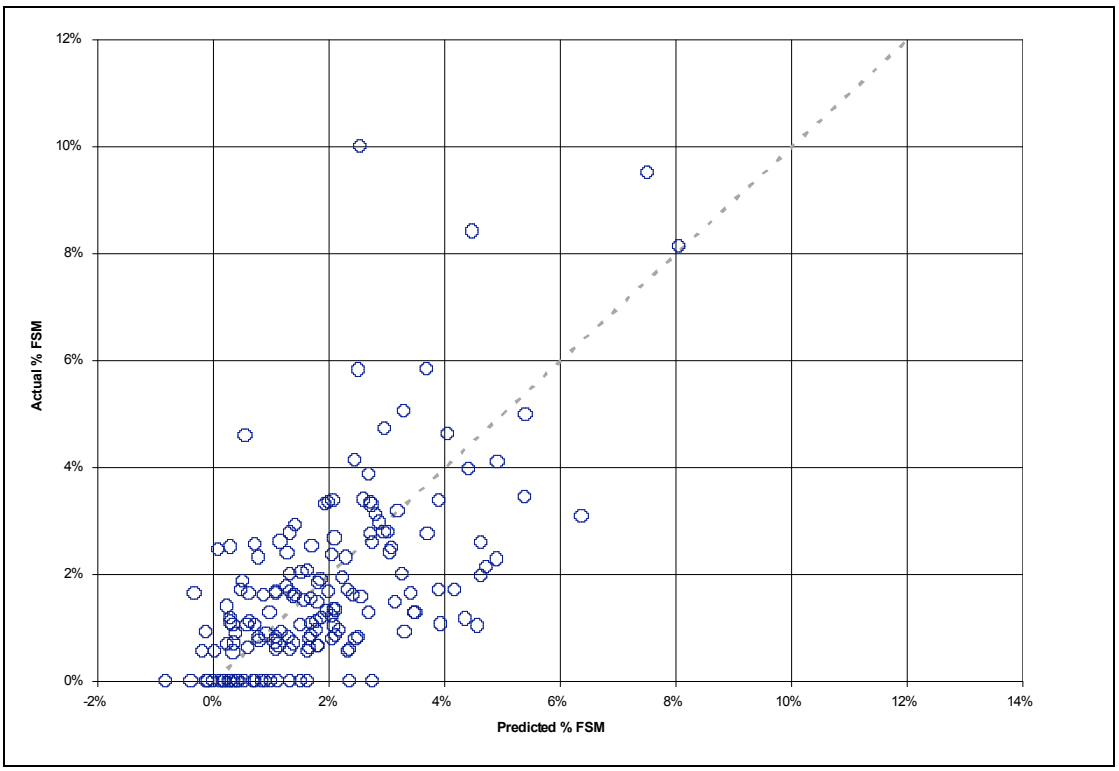
To answer this question a number of variables were entered into a regression model to predict a school's percentage FSM. The best fit came from including just two of these explanatory variables: the average KS2 score for pupils in the school and the percentage FSM for pupils in the catchment (i.e. in wards where pupils at the school live) who do not attend the school. With these two factors, the model explains about half the variance in the school's %FSM (Adjusted R-squared = 0.49; R=0.7[10]). Other variables, such as the level of selection in the LEA in which the school is based, or whether the school is in an urban or rural location, do not account for any further variation.

Figure 22 shows the actual percentage of pupils eligible for FSM in each of the 164 grammar schools, together with the percentage that is predicted by this model. The distance of each school above or below the line indicates the difference between its %FSM and what might have been expected. For example, a number of grammar schools have no FSM pupils at all (plotted along the horizontal 0% line). Some of these (towards the left hand side of the graph) would not be expected to take any FSM pupils, given the ability levels of their cohort and the prevalence of FSM in the locality. Others (towards the right) might be expected to have a cohort with up to 3% eligible for FSM. Further up the graph can be found grammar schools with some FSM pupils, either above the dotted line (indicating that they actually have more FSM pupils that might have been expected from their location and academic standards) or below the line (indicating they have fewer FSM pupils than expected). Almost all schools are within about 3% of the predicted value.

[10] The use of the logit-transformed percentages did not improve this fit.

Given that many grammar school cohorts are not large, such a small variation from expectation is probably not significant.

Figure 22: Grammar schools: Percentage FSM, compared with what would be expected given their ability profile and socioeconomic context



The hope of being able to identify some grammar schools that are taking more (or less) than their expected share of FSM pupils was unfortunately dealt a further blow by an analysis of the stability of these differences over time. The distance above or below the line for each school (its residual from the regression model) was calculated for the 2006 GCSE cohort and for the 2004 cohort. The correlation between these two residuals was just 0.23, suggesting low stability. Interestingly, the percentage of FSM eligibility among the neighbours of those in each cohort was very stable (r=0.93), as was the average KS2 score of those attending the school (r=0.91), especially considering the limited range in these variables. There was some variation in the actual percentage FSM of the pupils in each cohort (r=0.60).

7.2.9. *Can grammar schools' low rates of FSM be explained by their high academic standards?*

A final analysis of the social composition of grammar schools looks at the relationship between FSM and KS2 rather than the 'selectivity' of either.

Figure 23: Relationship between FSM status and KS2 scores for grammar and non-grammar school pupils, and for non-grammar pupils in grammar school areas

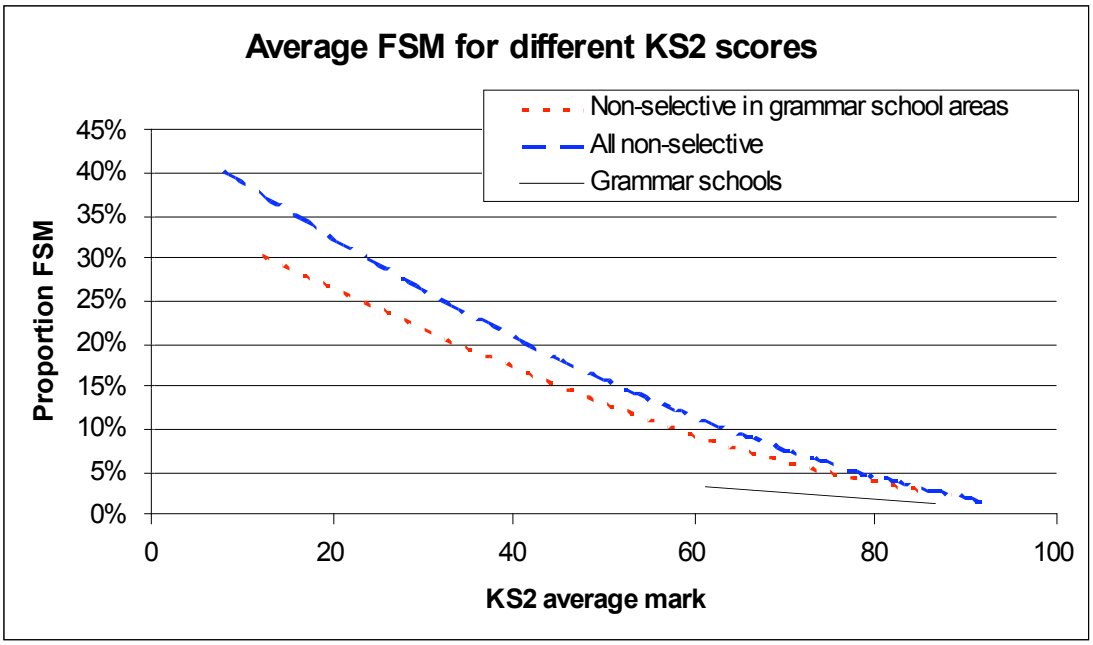


Figure 23 shows the relationship between FSM status and KS2 scores for both grammar and non-grammar school pupils. For each type of school, pupils were allocated to bins of at least 500, each with a narrow range of average KS2 scores. For each bin, the percentage of pupils eligible for FSM was calculated and plotted on the graph. A trend line was then fitted to these points.

It can be seen that for non-grammar school pupils, there is a steady decline in the incidence of FSM as KS2 scores increase (blue dotted line). The same is true for pupils in grammar schools, but for the same KS2 score these pupils seem to be only about half as likely to be eligible for FSM (black line). Of course, this may not be a strictly fair comparison, since many grammar schools are located in areas with relatively low FSM prevalence. However, limiting the non-grammar school population to those pupils in schools creamed by grammar schools (those that lose at least 5% to grammar schools) and hence in the same areas, lowers the line only slightly (red dotted line). The low rates of FSM eligibility found in grammar schools cannot, therefore, be explained by the academic ability of the pupils. Instead, there is a strong suggestion that some bias may be operating in the application or selection processes of grammar schools, which makes FSM pupils less likely to attend them – or makes those who do attend them less likely to be identified as eligible for FSM.

A possible alternative explanation for this phenomenon could be that reporting rates of FSM eligibility are lower in grammar schools than in other schools. A child who would be eligible for FSM who attended a grammar school might find themselves the only one, or one of only a handful, in that category in the school. They might feel stigmatised as part of such a small minority and choose not to draw attention to themselves by claiming their eligibility. Equally it might be that schools could differentially encourage the reporting of FSM status among their pupils. Schools with relatively low

academic achievement might work harder to have their rates of FSM recognised in order to make their performance seem more creditable. Those whose raw performance was high (eg grammar schools) might feel less need to emphasise their rates of FSM.

The PLASC dataset contained a marker for FSM for pupils in Year 9, but also for those same pupils when they were in primary school, in Year 6.[11] For the 582,000 pupils with valid FSM data on both sweeps there is a significant turnover, with 10% of pupils changing their status. Overall there is a small drop in the proportion eligible for FSM, from 17.7% in Y6 to 15.6% in Y9. The rates of FSM eligibility and the absolute changes seem quite different between those who attended grammar schools (from 3.50% to 3.02% eligible for FSM in Y6 and Y9, respectively) and those who did not (from 18.2% to 16.0%). Nevertheless, the odds ratio for these changes is very close to 1 (0.98), suggesting that both groups are dropping out of FSM eligibility at about the same rate, given the underlying prevalence. Hence the hypothesis that FSM is relatively under-reported in grammar schools seems not to be supported.

[11] These data relate to the 2007 GCSE cohort who were in Y6 in 2002 and Y9 in 2005.

8. Comparison of the performance at KS4 of selective and non-selective schools

It might seem like a simple matter to take a standard measure such as GCSE and compare the performance of pupils in grammar and non-grammar schools. However, the criticisms we have made of the over-simplistic approaches adopted by many previous studies (see Chapter 3) suggest that this is not the case. Indeed, we found that many of these criticisms were themselves so complex that we needed an additional chapter (Chapter 4) to explicate these issues. A comparison such as this is very far from being a simple matter; before conducting any analysis, we must think carefully about what may be the best way to do it.

One way to approach this complexity is to see the analysis process as consisting of a series of choices to be made. This is the approach we adopt. In essence we suggest there are four different kinds of choices that any would-be analyst must face. They concern how the different Key Stage 4 outcomes should be treated, what kinds of factors should be taken into account in order to make comparisons fair, what kinds of statistical models should be used, and which groups should be compared. Given that there are usually a number of options under each broad heading, and that each choice under one can generally be combined with all combinations of choices under the others, the number of permutations quickly becomes large. We estimate that there are over 500 defensible alternative ways of making the comparison.

We should also point out that even with this large number of alternatives, we have still had to make some assumptions that close down the possibilities. For example, the only outcome measures we consider are academic and vocational GCSE qualifications achieved at KS4 (aged 15-16). We could equally well have looked at KS3 as an outcome, or even A-level.

In the following section, 8.1, we outline some of the issues that arise in making these choices about how to proceed. In the subsequent section, 8.2, we present the results of making the comparison using different combinations of those choices.

8.1. Choices to be made in comparing performance

8.1.1. How should the different Key Stage 4 outcomes be treated?

Comparing the Aggregated Outcomes – is there a best way to measure it?

The total points score can be criticised as a representative measure of outcome since the more GCSE subjects a pupil sits the higher his or her points score is likely to be. If schools are judged on total points score this may encourage more examination entries per pupil than is educationally sound. In

our analysis we found over 6000 pupils entered for 12 or more full GCSE subjects, and of these about 1600 had taken at least one short GCSE as well. If all GCSEs and their equivalents, as deemed by the DCSF are included, then according to DCSF figures over 90000 pupils, or about 17.5% of the total entry, took 12 or more GCSE subjects and equivalents.

The average points score can also be criticised for failing to take account of the number of subjects entered. A candidate who enters only one subject would be judged to have achieved as much as another who achieves the same average grade on ten or more examinations. As with total points score, the average hides any differences in effort that pupils may have put into different subjects.

It can be argued that capping the number of GCSEs that count as the 8 best results is the fairest way to compare performances. However this could encourage schools to enter their pupils for the less demanding subjects. Also not all pupils take 8 GCSEs. In 2006 about 30% of pupils took 7 or less full GCSEs, but of these about half of them had taken at least one short GCSE. In terms of GCSE subjects and their equivalents, DCSF figures indicate about 10% of all pupils, or about 54000 pupils, were entered for 7 or less.

The above figures indicate the difficulties inherent in how best to make a comparison of achievement at KS4 of the grammar school population with the non-grammar school population.

We considered the following five aggregated KS4 achievement scores as output variables:

Total points score; all GCSEs and their equivalents

Capped points score; best 8; all GCSEs and their equivalents

Total points; full and short GCSEs

Average points score; full and short GCSEs

Rasch[12] score of overall GCSE achievement

Individual GCSE subject points scores:

Mathematics	English
Double science	History
English literature	French

These subjects were chosen as a mix of the subjects which are compulsory at KS4 and typical choices made in year 10 in a grammar school curriculum. We did not consider vocational GCSEs in the individual subjects.

8.1.2. What kinds of factors should be taken into account in order to make comparisons fair?

Raw GCSE results can be criticised as a way of comparing individual pupil performance in that it is a one-off snap-shot of what he or she achieved in a set of examinations sat at one time (or largely at one time). As such there is no measure of relative progress between pupils relative to some defined starting point. Value added analysis addresses this criticism by measuring

[12] The Rasch score is a technique which takes the difficulty of the subject into account. See p145.

performance relative to a starting point usually said to be a *baseline*. The baseline might be the pupil's achievement at Key Stage 2 or 3, but how should this achievement be measured? The measures available are the levels awarded following formal tests, the actual scores in mathematics, English and science, and the teachers' assessment of the level of the pupil on the National Curriculum scale. Other measures might be used as in the CEM Centre's tests of developed ability, which pupils can take at various stages in their primary and secondary education.

In this section we do not consider Key Stage 3, but rather consider progress over the 5 year period from entering secondary school in Year 7 to taking GCSE examinations in Year 11. This 5 year period is chosen since the important question behind this whole study is to what extent can achievement or measured potential at age 10/11 predict educational outcome as measured by examination results 5 years later? This is the question that selection at age 11 for a place at a grammar school is all about. However, it should be noted that for various reasons, not all pupils who have a performance record at KS4, have a recorded performance record at KS2 from 5 years earlier. In our analysis we included only those pupils who had a complete KS2 record of a level in each of mathematics, English and science; that is a total of 533883 pupils.

A fairly simple question to ask is, 'given what we know about a child at age 10/11 can we predict his or her likely performance in examinations at age 16?' We recognise that that there are many factors that could have some bearing on a child's development during secondary school, other than his or her so-called innate ability. We will not enter into a discussion on any differences between ability, aptitude and interest here except to say that a pupil's attitudes to school and the various school subjects and activities will have a bearing on their development. Other factors such as home background and parental attitude to education will be influencing factors, as will the general ethos of the secondary school attended, its location and its ethnic and social mix of pupils.

So in trying to answer the question about how achievement at Key Stage 4 relates to achievement at Key Stage 2, we had to decide what other baseline factors (or variables) we needed to control for. That is to try to recognise and measure their effect in accounting for the variability found between pupils in their Key Stage 4 performance who on Key Stage 2 performance alone might be deemed to be similar.

In practice there are at least three possible ways one could make the decision about what explanatory variables to include in a statistical model. Firstly, a common approach adopted by statisticians would be to apply statistical criteria. If a particular variable, in conjunction with others in the model, accounts for statistically significant variation in the outcome measure, then we should include it, otherwise not. Secondly, we could apply theoretical criteria. If we have reason to believe that an unfair comparison would arise from failing to take account of the effect of a particular variable, then we should include it. Thirdly, we are inevitably constrained by pragmatic criteria. If we have a measure of a particular variable available, then we may include it, otherwise we cannot.

After some initial analysis, and combination of these three approaches, we chose variables at two levels: pupil-level and school-level.

At the pupil level, prior attainment measures in the form of Key Stage 2 scores were available in the dataset and these seem to be the obvious choice for the main explanatory variable. However, even here there are a number of choices to be faced.

Most previous studies have entered levels for each of the three subjects (maths, English and science) separately (e.g. Schagen and Schagen, 2005). No study that we are aware of has used the actual marks achieved, which is perhaps surprising given the arguments outlined in 4.2 (p112). However, there are some problems with the use of marks, in particular the fact that marks do not correspond with levels. Although the majority of pupils with a given level have marks within a narrow range, a significant minority fall outside – sometimes well outside – this range. This lack of correspondence is one reason the DCSF (and previously the DfES) make no official use of marks.

When considering the use of marks it became evident from the examination of their distributions that they are substantially skewed (see Figure 2, p110). For this reason, we considered the average normalised mark from the three KS2 subjects, though we found the model fit was improved only very slightly. We therefore used KS2 marks rather than normalised KS2 marks in our analysis.

Similarly we considered entering the three KS2 subjects mathematics, English and science, separately into the model. Although this did improve the fit slightly, this gain was not felt to outweigh the extra complexity of having additional variables in the model and possible problems of co-linearity, though this decision was perhaps somewhat arbitrary and might have made a small difference to the results. Two variables are co-linear if they are highly correlated, and this creates problems in interpreting the results of a regression analysis.

When considering the use of KS2 levels, we again considered average levels, rather than the individual subjects, for the same reasons given above. We decided to only include students who had results for all three KS2 subjects and eliminate the minority of pupils who had results for two or less subjects. This was because preliminary regression analysis showed that these minority groups of students behaved differently to those who had results for all three subjects and therefore their inclusion could be problematic. The reasons for these differences could be many and varied. As so few students in grammar schools had missing KS2 data, the simplest option was to remove them from our main analysis.

We also considered including the KS2 level as assessed by a pupil's teacher as well as the level achieved by the pupil through the KS2 assessment tests. However, this led to problems of co-linearity, the teacher assessment and the test assessment being closely correlated, and ultimately the teacher assessed level variable was dropped from the analysis.

A further problem that arose was the question of how to treat ethnicity in this kind of analysis. This question is far from simple.

Within the PLASC dataset, pupils are classified into 20 different ethnic categories. One approach would be to include all 20 separately in the analysis. However, one could certainly argue that ethnicity interacts with other variables in influencing the progress that pupils make, so these interactions should also be included. For example, girls of a particular ethnic group may

be quite different from boys, and it could be that the effect of ethnicity depends also on prior attainment. To include even just these two possible interactions would triple the number of explanatory variables in the model, requiring 60 variables just to estimate the impact of ethnicity. Clearly such a proliferation would be absurd, especially given the small numbers of pupils in some of the categories.

In fact, if we look at the numbers in each ethnic group in the DCSF supplied PLASC data, we see that only four of the categories contain as much as 2% of the total population. Of these, the category 'white British' is totally dominant at about 70% of the population, followed by 'information not obtained' at about 15%, the other two categories being 'Indian' and 'Pakistani' at 2% each. Hence, one must presumably combine some of the categories before they can be used in the analysis, but it is far from obvious which ones should be combined. For example, combining all the 'other white' categories accounts for 3.4% of the data.

Fortunately, the situation is simplified a little by the fact that our purpose here is not to estimate the impact of different ethnic classifications, but to estimate the impact of grammar school education. If the proportions in each ethnic group are similar in grammar schools to those in non-grammar schools, then there is little point in including a main effect for membership of that group, since it will have no effect on any estimate of the difference.

Table 39 shows the number and proportions of the eight most dominant ethnic categories that we initially decided to work with, in our population of 533883 pupils.

Table 39: Largest ethnic categories

Ethnicity category	grammar schools			non-grammar schools	
	Number	Percent		Number	Percent
White British	17125	81.1%		427629	83.4%
Indian	980	4.6%		10847	2.1%
Other White	556	2.6%		10151	2.0%
Pakistani	331	1.6%		11049	2.2%
Chinese	249	1.2%		1373	0.3%
Caribbean	111	0.5%		6934	1.4%
African	110	0.5%		5751	1.1%
Bangladeshi	63	0.3%		4778	0.9%
Other	1530	7.2		31719	6.2%
not obtained	67	0.3		2530	0.5%
Total	21122	100.0%		512761	100.0%

In Table 39 it can be seen that although the proportions in the grammar schools and non-grammar schools are similar, in the ethnic minorities it can be seen that there are proportionally more Indian and Chinese pupils in the grammar schools, and proportionally more Pakistani, Caribbean, African and Bangladeshi pupils in the non-grammar schools. The "other" category includes pupils from any other Asian or black background, and any other ethnic group.

In our statistical analysis we attempted to investigate the extent of the effect of ethnicity using these categories. However in some of the statistical models considered problems arose. Retaining the individual ethnic groupings as in Table 39 led to numerical instability in some of the models. This is probably due to the relatively small numbers of pupils in most of the categories. Ultimately in order to achieve well behaved statistical models the ethnic categories were regrouped into three categories A, B and C these being A, the "other" category from Table 39, B, all other known ethnic categories combined and C, white.

A further problem we needed to consider was how to take into account a pupil's socio-economic status (SES). An apparently simple measure of SES is a pupil's eligibility to free schools meals (FSM). However, the situation is not completely clear as not all those pupils, or rather their parents or those legally responsible for the pupil, claim FSM even if eligible by way of income due possibly to the social stigma attached. The National Census Data (2001) provides a range of indices which are used to measure deprivation at ward level within a local authority. One of these indices, the income deprivation affecting children index (IDACI) was available in the 2006 PLASC data and so we have included this index and FSM in some of our models.

In the end, our choice of individual pupil-level variables was:

Key Stage 2 achievement in mathematics, English and science:

1. The average level as determined by the KS2 tests.
2. The average mark actually obtained on these tests

It was not clear at the outset as to whether either of these measures of KS2 achievement would be more reliable than any other, thus we included both of them in the statistical models we considered.

Factors relating to the social aspects of an individual pupil:

3. Sex
4. Ethnicity as categories A, B and C
5. Deprivation indices; FSM and IDACI.

And, most importantly for this study:-

6. Whether or not the child attended a grammar school (more precisely was attending a grammar school at the time they took their GCSE examinations in 2006)

When we come to consider school-level variables, the issues are particularly complex. An especially difficult dilemma arises in relation to whether a variable should be included or not in the case of 'compositional' variables. These variables indicate characteristics of the composition of a body of students in a school such as their average ability or socio-economic status. They are typically formed by aggregating individual student-level variables, which may themselves also be entered in the model in their own right. Many school effectiveness studies have found that such aggregated school compositional variables explain variation in the outcome measure above and beyond what is explained by the same variables at the individual level. When this happens, it is often referred to as a 'compositional effect' (Harker and Tymms, 2004). However, other studies have failed to find such effects (see

Nash, 2003; Teddlie, Stringfield and Reynolds, 2000) and they remain controversial within the field of school effectiveness research.

Where there is a statistical compositional effect the obvious interpretation is that two students who are the same in every respect apart from the school they attend will be expected to make different progress. Usually, the one who attends a school in which the average level of achievement – or social advantage – is higher will make more progress, though occasionally 'negative' compositional effects have been found, reversing this tendency (Harker and Tymms, 2004).

The problem is that we do not know whether this relationship is because more-advantaged schools also tend to be genuinely better schools, or whether this appearance is misleading and results from a failure to take account of other differences. In the former case we should certainly not include the compositional variable in the model, since to do so would be to eliminate a real difference. For example, it could be that schools with more able populations of pupils are also generally able to attract better teachers and thus do genuinely provide a better education and enhanced performance. If we include the average ability of the cohort in our model as an explanatory variable, we will automatically level the playing field and make schools with different ability profiles systematically equal in their performance.

On the other hand, it could equally be that the tendency for schools with high ability intakes to appear to perform better is because of other unmeasured differences in the types of pupil they attract. In this case it would be wrong not to try to adjust for these differences and hence we should include the compositional variable in the model, since not to do so would be to bias the results unfairly against schools with the most disadvantaged intakes.

Unfortunately, it is not easy to know which of these cases prevails. It seems likely that they are both true to some extent; the problem is where to strike the balance. Our approach to this problem was to run the statistical models both with and without compositional variables included. The two results may then be seen as providing upper and lower bounds for the 'true' grammar school effect.

The final choice of compositional variables relating to the secondary school as a whole was:

1. Average Key Stage 2 level in a year group cohort.

We assumed that average level would be a representative measure if taken across a whole year group, so did not consider average mark here.

2. The proportion of pupils in the school eligible for free school meals. As the relationship is non-linear, we also included the squared term. Also the average IDACI index for the school.
3. Whether the school is single sex (for age 11-16) or coeducational.
4. If a single sex school, whether it is a boys' school or a girls' school.

8.1.3. What kinds of statistical models should be used?

We used two statistical models for investigating the value added for each pupil between Key Stage 2 and their achievement at KS4. These were:

Ordinary Least Squares regression (OLS)

Multilevel (ML) Modelling

We will not go into any detail of the mathematics that lies behind these statistical models, but will give in broad outline a description of the technique of regression on which both models are based. Regression is based on the assumption that there is a relationship between an input variable (for example average Key Stage 2 level) and an output variable (for example, total point score at KS4), and the technique of linear regression establishes a best fit as a straight line graph for that relationship, amongst the scatter that inevitably exists in the raw data. The best fit is obtained by minimising the sum of the squares of the residuals between the line and the actual data points, where the residual is the vertical distance from the line to the actual data points. Thus the phrase *least squares* regression.

The input variable in regression analysis is often called the explanatory variable, in that this variable is used to explain at least some of the behaviour observed in the output variable. In a complex situation such as that found in education, and many of the social sciences, there are often many possible input, or explanatory, variables and also many possible output variables. This gives rise to multiple regression. In multiple regression the input variables are input to the model in varying combinations. We carried out several multiple regression analyses. What we were seeking is the *grammar school effect*. That is, what difference to the output variable does attending a grammar school make. This was measured in terms of GCSE grades. Thus our prime interest was in the input variable (6) above; whether or not the pupil attended a grammar school. By inputting other variables as well we *control* for their effect on the output variable, in that the model lets each input variable account for some of the variation in the output variable. There is a constraint on this in that all the regression best fit lines are assumed to have the same slope. This constraint can be removed in a multilevel model.

Multilevel modelling is a relatively recent (from about 1980) development of multiple regression (Goldstein, 1995). The essential difference from OLS multiple regression is that a multilevel (ML) model recognises the fact that pupils are grouped within schools in a way that OLS does not. In an OLS model, all pupils are treated the same (and assumed to be independent), regardless of which school they attend. The pupil-level results (residuals) are then aggregated to the school level in order to draw conclusions about the school.

In a multilevel model, the relationships are essentially modelled within each school and across all schools simultaneously, allowing the components of variance to be partitioned as 'pupil-level' (i.e. the amount pupils vary within the same school) and 'school-level' (the variation across schools). Multilevel modelling has come to be seen as the orthodox approach for studies of school effectiveness.

Within a multilevel model there is a further option which is to allow the slopes to vary across schools. Although a ML model with fixed slopes effectively estimates a separate regression line for each school, all these lines are parallel to each other (the slope is fixed). In a 'variable slopes' model, on the other hand, the slopes of individual school regression lines can vary. This means that it is possible to identify not just whether one school is more effective than another, but to say that one may be more effective with able pupils, while another may be better with those of low ability; others may be good with all, or with neither. Although this seems like a substantial advance

it brings with it the complication that we can no longer summarise the effectiveness of a school, or group of schools, with a single number, since 'effectiveness' will vary for different pupils.

8.1.4. Which groups should be compared?

An obvious starting point for comparison groups is to compare the performance of pupils in grammar schools to those in other schools (comparison group number 1 below). However we also need to consider the effect of selective systems as a whole, by comparing selective systems (grammar schools *and* the schools they cream from) with non-selective systems to investigate whether grammar schools are having an effect on the schools that they cream from. This is not as easy as it seems. We have shown in Section 6.3.2 (p143) that the amount of boundary crossing between LAs makes it meaningless to use the LA as a unit of measurement (however we have included this here as comparison group number 5 as this is the unit that has been used in many studies to date). In Section 7.1.7 (p160) we attempted to use factor analysis to define clusters that make up selective systems, but no clear groups were found. We could use the official designations of grammar, secondary modern and comprehensive from the DCFS (formerly DfES) (comparison group number 6) but these designations seem little more than arbitrary as schools designated as comprehensives can be equally close geographically to grammar schools and as equally creamed as secondary moderns. In fact many LAs contain both types of school. We are therefore left with a decision about what percentage creaming to use to define a selective system. This decision is essentially arbitrary – 20% would indicate schools being heavily affected by the grammar schools, 5% could be equally argued as a good choice and 1% is probably the least detectable level of creaming, so we have chosen to include all three cut-off points (groups 2 – 4):

1. Grammar schools vs all other schools
2. Grammar school and >20% creamed vs all others
3. Grammar schools and >5% creamed vs all others
4. Grammar schools and >1% creamed vs all others
5. Selective LAs (i.e. at least 10% in grammar schools) vs others
6. Grammar schools and secondary moderns vs comprehensives

8.2. Direct comparisons

The Key Stage 4 GCSE results analysed and considered in this and the following sections are for 533883 pupils who sat their GCSE examinations in summer 2006. This is the number of pupils who we were able to match with their Key Stage 2 results from when they were in Year 6 in 2000-01 and had an assessment test based level recorded for each of mathematics, English and science. Of these pupils, 21,122, or 3.96% of the cohort analysed, took their GCSE examinations at a grammar school.

8.2.1. Raw results for individual GCSE subjects

In this initial analysis we discuss the GCSE results for the three compulsory National Curriculum subjects: mathematics, English and science. Initially we consider only the double science GCSE as this is the only option chosen by the

vast majority of schools, as opposed to the single science GCSE, or GCSE in the actual sciences of biology, chemistry and physics.

Figure 24: Distributions of grades awarded in Maths, English and Science

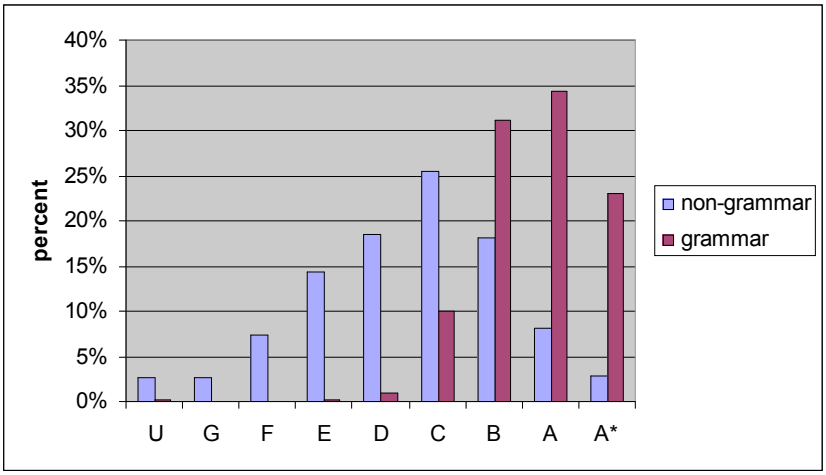
Mathematics

Number of grammar school entries 21,033

(99.6% of the cohort)

Number of non-grammar school entries 498,490

(97.2% of the cohort)



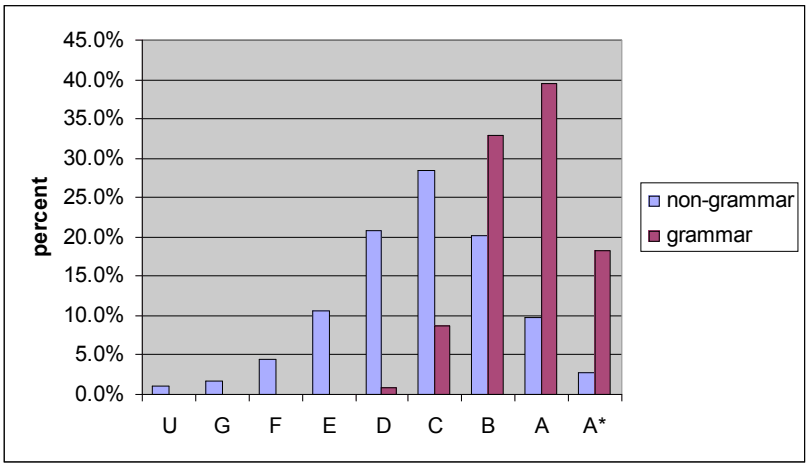
English

Number of grammar school entries 21,062

(99.7% of the cohort)

Number of non-grammar school entries 494,894

(96.5% of the cohort)



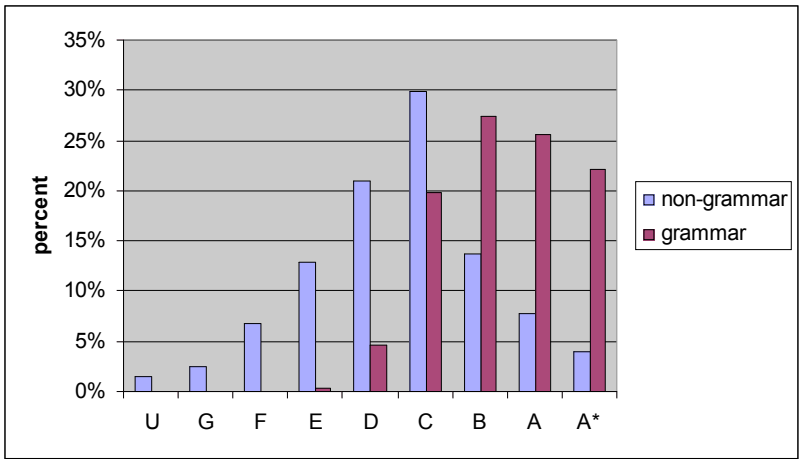
Double Science

Number of grammar school entries 13,679

(64.8% of the cohort)

Number of non-grammar school entries 372,881

(72.7% of the cohort)



In Figure 24 we present the distribution of the grades awarded to the pupils, comparing the results of pupils who attended grammar schools with those who did not. We refer to these schools as non-grammar schools rather than use the blurred distinction between comprehensive and secondary modern schools.

The grammar school entry for double science is relatively low compared with mathematics and English, as a large number of grammar school pupils were entered for separate science examinations in biology, chemistry and physics. A more detailed analysis of the GCSE entry to science subjects is shown below, in Figure 25.

The results in Figure 24, as might be expected, show a much higher performance by grammar school pupils in terms of percent of the entry cohorts achieving the higher grades. In mathematics 57.5% of the grammar school entry achieved a grade A or A*, and 57.7% in English. This compares with 11.1% and 12.6% respectively for the non-grammar school entries. Similarly in mathematics for the grammar school entry only 11.3% of the entry were awarded a grade of C or below compared to 70.8% in the non-grammar school entry. The corresponding figures for English are 9.5% and 67.3 % respectively.

Figure 25: Distribution of the grades awarded in Science subjects

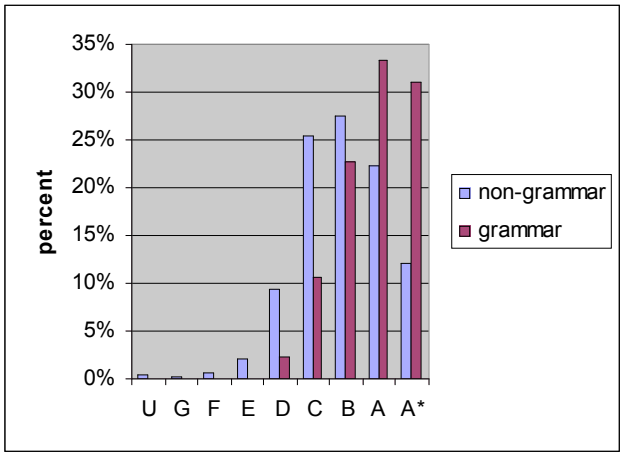
Double Science

Number of grammar school entries 13,679

(64.8% of the cohort)

Number of non-grammar school entries 372,881

(72.7% of the cohort)



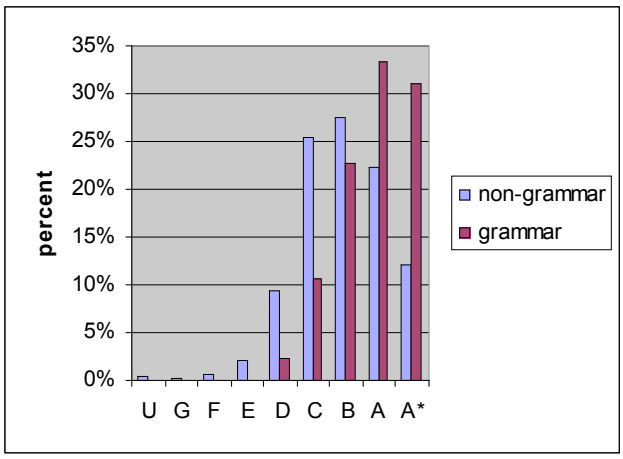
Physics

Number of grammar school entries 6,876

(32.6% of the cohort)

Number of non-grammar school entries 25,850

(5.0% of the cohort)



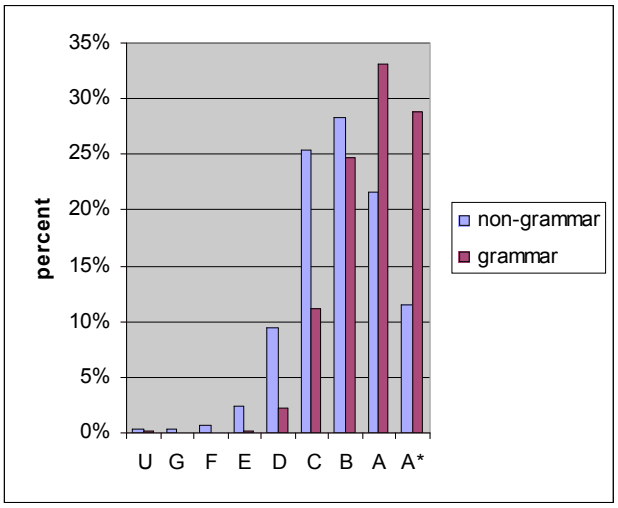
Chemistry

Number of grammar school entries 6,880

(32.6% of the cohort)

Number of non-grammar school entries 26,090

(5.1% of the cohort)



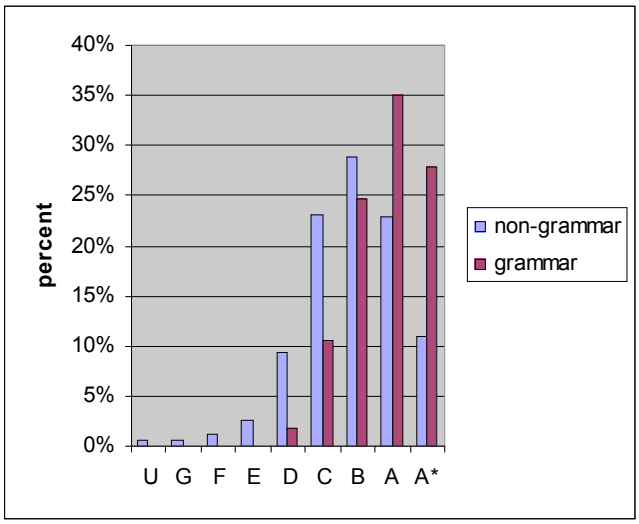
Biology

Number of grammar school entries 6,893

(32.6% of the cohort)

Number of non-grammar school entries 27,353

(5.3% of the cohort)



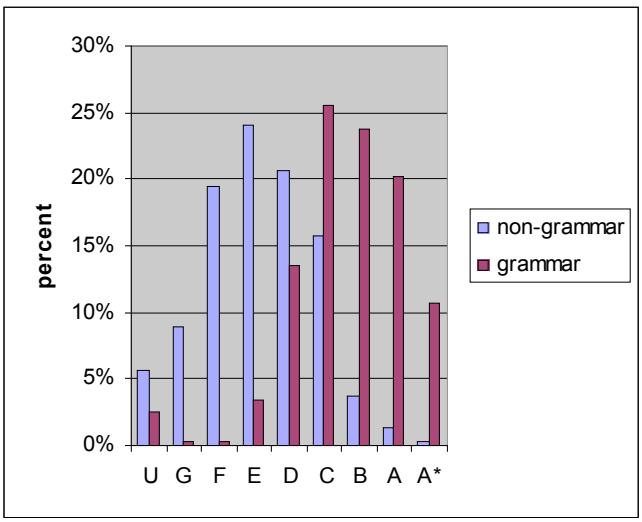
Single Science

Number of grammar school entries 357

(1.7% of the cohort)

Number of non-grammar school entries 55,460

(10.8% of the cohort)



The most notable difference for the separate sciences are that the entry numbers are far smaller than for double science, although it is notable that about one third of the grammar school pupils are entered for these; presumably these are mostly the same pupils taking three separate sciences although this cannot be confirmed from this current analysis. In contrast

about 5% of the pupils in the non-grammar schools were entered for the separate sciences, and their achievement is notably lower than that of the grammar school pupils. Whereas over 60% of grammar school pupils achieved a grade A or A* in all the three sciences only about 30% of non-grammar schools pupils achieved this. Nearly all the grammar school pupils, about 98%, achieved a pass grade compared to around 85% for the non-grammar school pupils.

The single science GCSE is intended for those pupils considered not capable of taking the double award, so it is surprising there were any such pupils in the grammar schools, although we see at 1.7% of the cohort, the entry was very small. A considerably higher entry would be expected from the non-grammar schools and we see this was about 11% of the cohort, and of these about 80% failed to achieve a grade C or above, compared to about 20% in the grammar schools although the actual number of such pupils is very small.

8.2.2. Aggregated GCSE results

As pupils at KS4 can take different types of qualification and achieve in each qualification at different grades or levels, the (then) DfES devised a way of aggregating GCSEs, GNVQs and other Level 2 qualifications results using a points system. These points when aggregated for each pupil are a measure of his or her overall achievement at KS4. Initially for GCSE this points system was 7 points for a grade A, 6 for a grade B and so on with 0 for a grade U. With the introduction of the A* grade at GCSE, and more vocational qualifications at level 2, this system was revised in 2004 so that grade A* is 58 points, grade A is 52 going down in steps of 6 so that a grade G is 16 points. The DfES made this revision so that all level 2 qualifications could be fitted in to the system with a deemed equivalence to GCSE at various levels of achievement. The total point score also included entry level and key skills qualifications, and GCE AS or VCE AS qualifications taken in year 11 or earlier. In 2006 the DfES had essentially two ways of aggregating the points scores. These are:

Total points score; all GCSEs and their equivalents

Capped points score; best 8; all GCSEs and their equivalents

As discussed in Section 8.1.1, as well as using these measures of achievement in our analysis, we have also devised our own measures. These are:

Total points; academic and vocational full GCSEs and short GCSEs

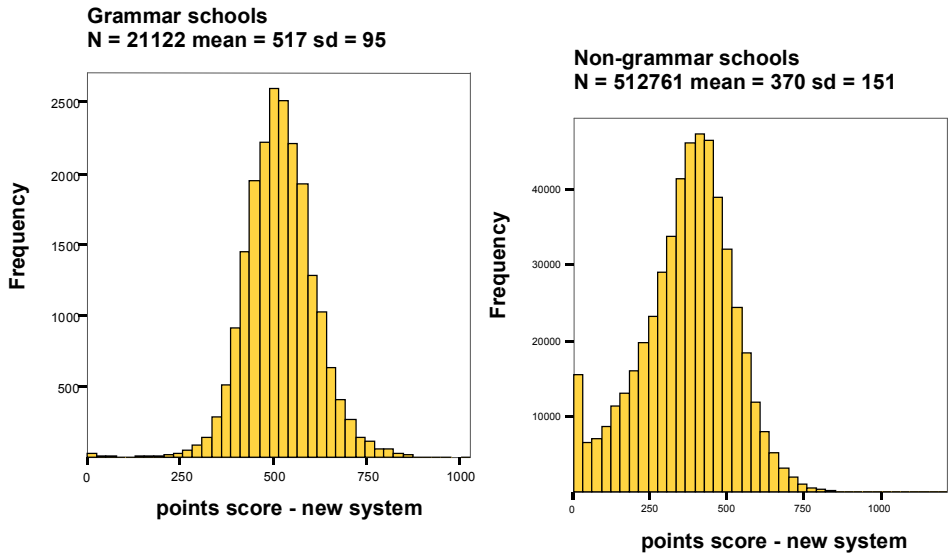
Average points score; the weighted[13] mean of the total points

Rasch score of overall GCSE achievement

The graphs shown in Figure 26 to Figure 30 below show these five ways of giving the aggregated points scores for grammar schools and for non-grammar schools.

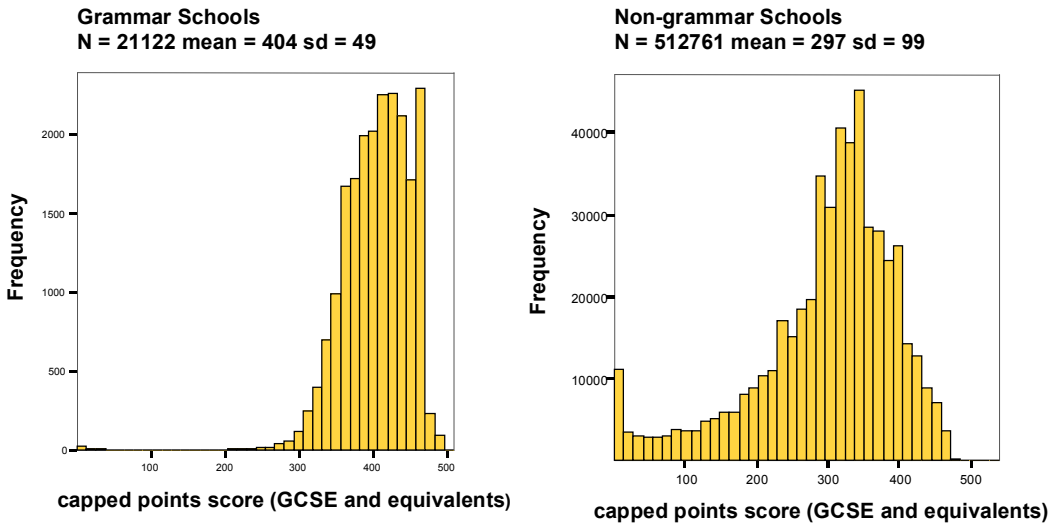
[13] Academic full GCSEs are weighted 1, except for double science which is weighted 2. Vocational GCSEs are weighted 2; short GCSEs are weighted 0.5

Figure 26: DCFS GCSE and equivalent Total Points Score for GS and non-GS



Grammar schools have a higher mean points score and a smaller standard deviation indicating a generally higher level of achievement and more consistent results across the schools. The grammar school points scores are also symmetrically placed about the mean values whereas the non-grammar points scores are skewed to the lower values. In the first category of Figure 26, that is less than about 30 points, there are about 15,500 pupils from the non-grammar schools, compared to 45 from the grammar schools.

Figure 27: DCSF Capped Points Score; for GS and non-GS

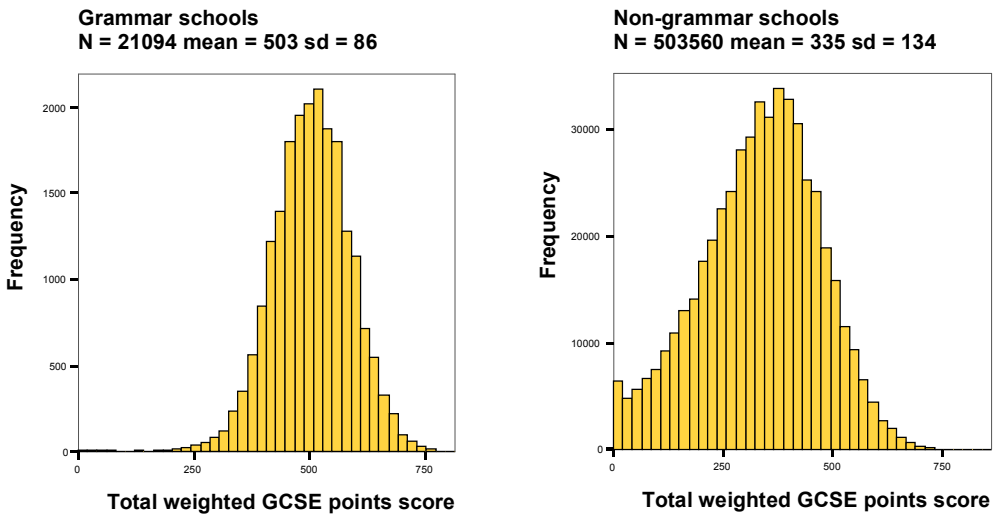


The capped point scores for the grammar schools indicate a ceiling effect. If the best score a pupil can achieve is 8 grade A*s then his or her capped score would be 464 points; although 1475 grammar school pupils achieved this total (about 7% of the cohort), the DCSF data also indicates 443 pupils with a capped score above this to a highest points score of 508, so just how the capped calculation has been done isn't clear. However, we can use it for

comparison purposes as the calculation is the same for all pupils. The mean value for the grammar schools is 107 points higher than that of the non-grammar school which equates to approximately two grade As at GCSE. The standard deviation in the grammar school point scores is about half that of the non-grammar schools, indicating again a much more consistent performance by the grammar school pupils. Very few of the grammar school pupils, less than 2%, scored below 300 points compared to a skew towards the lower points score in the non–grammar schools, where about 40% of the pupils scored below 300 points. However, there is comparable performance in the non-grammar school pupils at the high scoring end of the distribution; there were 2637 pupils (about 0.5% of the cohort) at the 8 A* points score level of 464, with a further 827 pupils above this with three pupils achieving the highest score of 540 points.

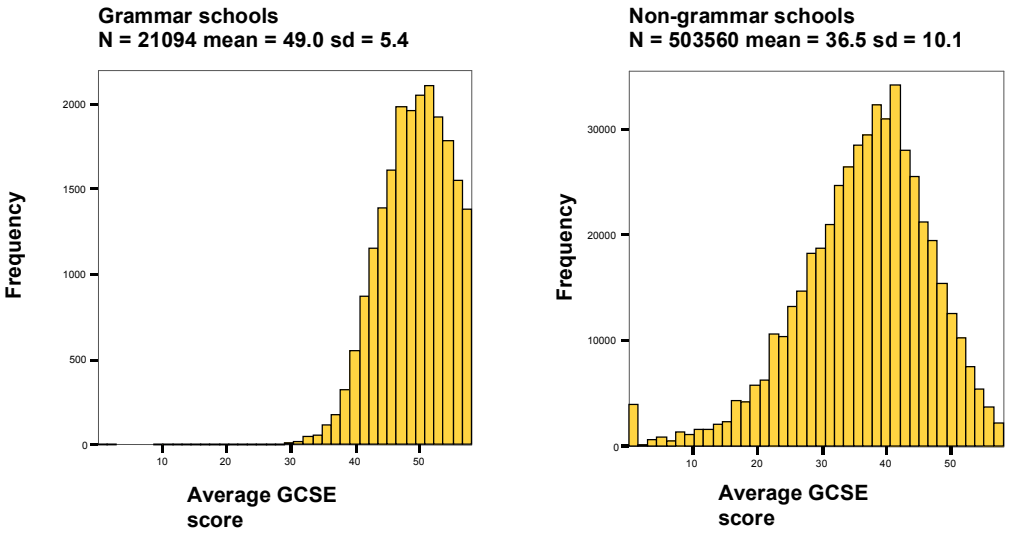
Due to the uncertainty in just how the DCSF has calculated the total points score and the capped scores, we have calculated our own scores based on the grades for full academic and vocational GCSEs and short GCSEs, using data as supplied by the DCSF. The distribution of the total points based on these qualifications is shown in Figure 28.

Figure 28: Total weighted GCSE Points Score



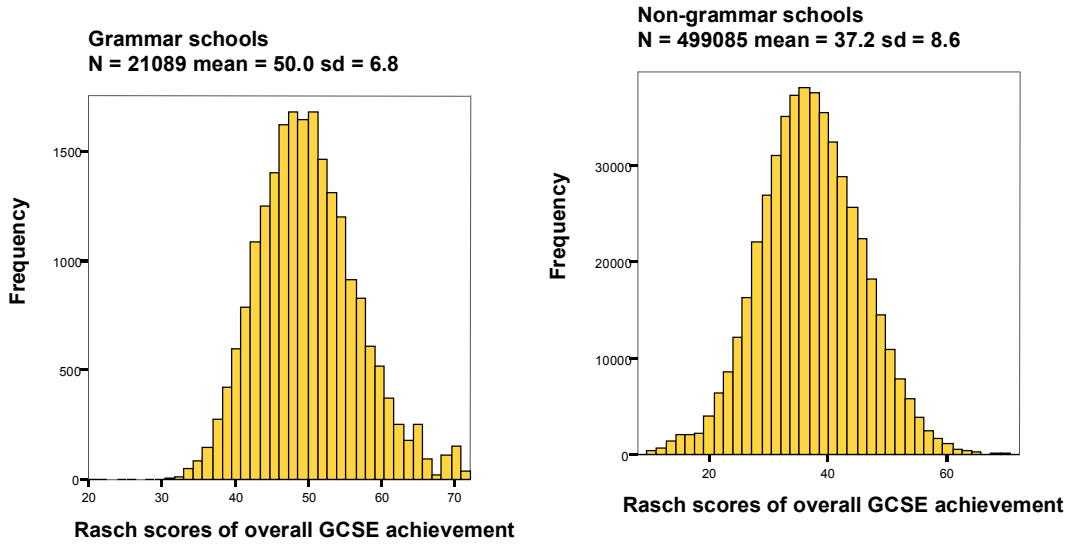
The two distributions shown in Figure 28 are very similar to those of Figure 26, using the DCSF point scores. It is notable that the cohort sizes have reduced slightly indicating that even in the grammar schools some pupils did not take a GCSE qualification. In the non-grammar schools there are notably fewer pupils who achieved at 30 points or less, but this would be expected as many low achieving pupils would not have taken any GCSEs. A wider spread is again seen in the non-grammar school point score both visually and in the standard deviation, but the two distributions are fairly symmetrically placed about their respective mean values.

Figure 29: Average Points Score



In Figure 29 a ceiling affect is again seen in the average GCSE scores for the grammar school pupils. There were 304 pupils actually at the maximum value of 58 with about 30% of the cohort averaging 52 (a grade A) or more points and about 70% averaging 46 (a grade B) or more. The mean value of 49.0 lies midway between a grade A and a grade B. The mean number of subjects taken by the grammar school pupils was 10.2 with a standard deviation of 1.2. The distribution for the non-grammar school pupils is more symmetrical about a mean of 36.4, which lies between a grade D and grade C. For the non-grammar school pupils there were 453 at the maximum of 58, with about 4% of the cohort averaging grade A and above, and about 16% averaging grade B and above. The mean number of GCSE subjects taken by the non-grammar school pupils was 8.7 with a standard deviation of 2.4.

Figure 30: Rasch scores of overall GCSE achievement



The Rasch technique used in this analysis takes into account the relative difficulty of the various subjects. Within this analysis the most difficult subjects have been identified to be the separate sciences of physics, chemistry and biology and foreign languages. Double science is relatively more difficult than most GCSE subjects but easier than the separate sciences. This probably accounts for the relatively high number of grammar schools who are achieving highly at the top of the distribution; it was noted earlier (Section 8.2.1) that about one third of the grammar school cohort were entered for the separate sciences and the majority of these achieved a grade B or better. Otherwise the graphs in Figure 30 are similar in shape to those of Figure 28, and again it is notable that the mean is higher for the grammar school pupils whilst the standard deviation is lower.

8.2.3. Matching by Key Stage 2 Level: Comparing Subject Grades

The graphs presented in 8.2.1 and 8.2.2 show higher achievement as a whole for pupils educated in grammar schools. However this is to be expected given the selective nature of grammar schools and the comparison could be criticised as not comparing like with like. At the high achievement end of the performance scale, there are pupils in the non-grammar schools whose achievement is at least as good as those in the grammar schools. Grammar schools select their pupils through entrance tests which are regarded as measures of ability and potential for later achievement. A direct comparison with non-grammar school pupils is not possible as these pupils do not take an entrance test.

However, if we regard three level 5s at KS2 as a measure of ability that is equitable to grammar school entry, then we have what could be argued as a fairer means of comparison. Children who gained a level 5 or higher in mathematics, English and science at KS2 at least demonstrated equitable achievement at that stage.

Here we present the graphs again but with the sample limited to those pupils who had achieved level 5 or higher in each of mathematics, English and science at KS2.

Figure 31: Distributions of GCSE grades awarded in Maths, English and Science for pupils with 5,5,5 at KS2

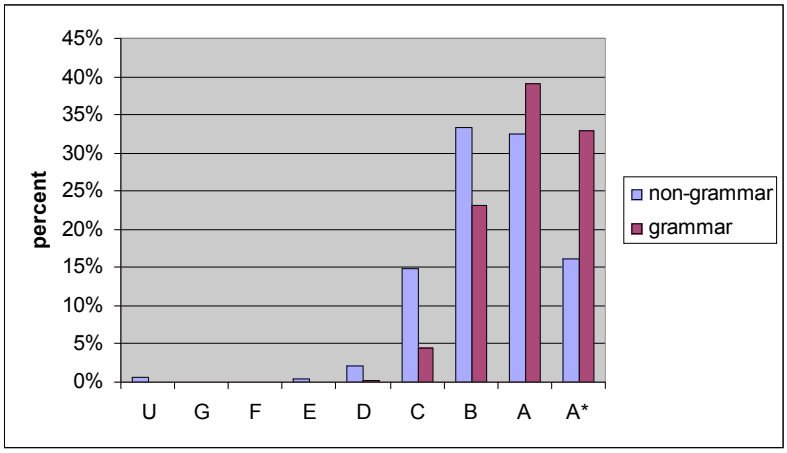
Mathematics

Number of grammar school entries 12,644

(59.9% of the cohort)

Number of non-grammar school entries 67,790

(13.2% of the cohort)



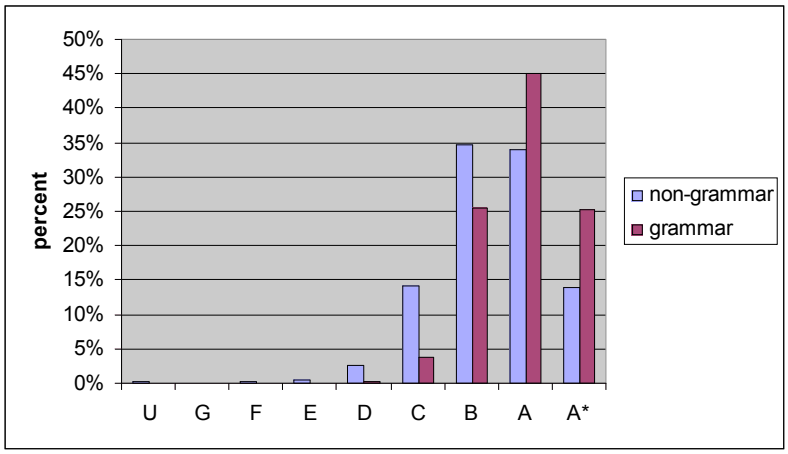
English

Number of grammar school entries 12666

(60.0% of the cohort)

Number of non-grammar school entries 67,783

(13.2% of the cohort)



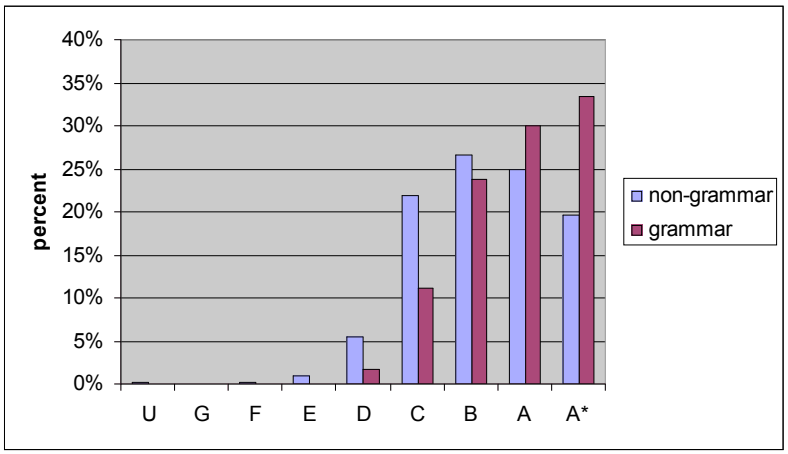
Double Science

Number of grammar school entries 7,420

(35% of the cohort)

Number of non-grammar school entries 51,918

(10.1% of the cohort)



In all three subjects the grammar school pupils are still seen to be achieving higher at GCSE than those from the non-grammar schools. In mathematics the number of pupils achieving grade A* or A is 72% of the entry in grammar schools, compared to 49% of the entry for non-grammar schools. Similar results are seen in both English and double science with corresponding figures of 70% and 48% for English and 63% and 45% for double science. In mathematics about 5% of the grammar school pupils achieved at grade C or below compared to about 17% for the non-grammar schools. The corresponding figures are 4% and 17.5% for English and 13% and 29% for double science.

In Figure 32 we present the grade distributions for the separate sciences of physics, chemistry and biology and also single science. As all pupils had three level 5's or higher at KS2, it might be expected that entry levels would be proportionately higher, although entry would depend on a particular school's entry policy.

Figure 32: Science grade distributions

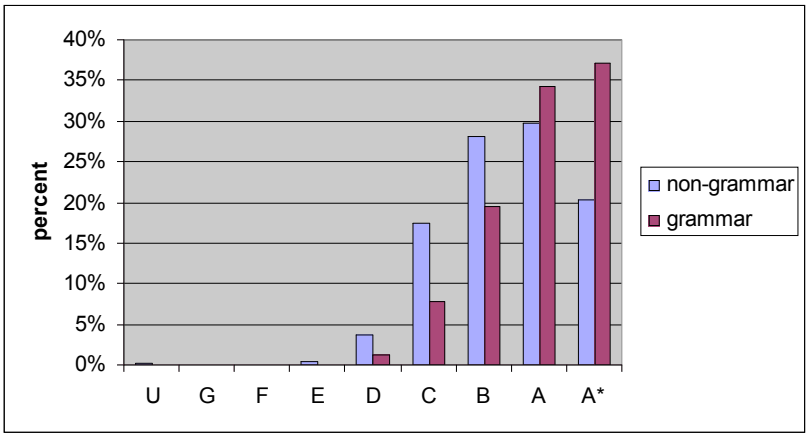
Physics

Number of grammar school entries 5,001

(23.7% of the cohort)

Number of non-grammar school entries 12,219

(2.4% of the cohort)



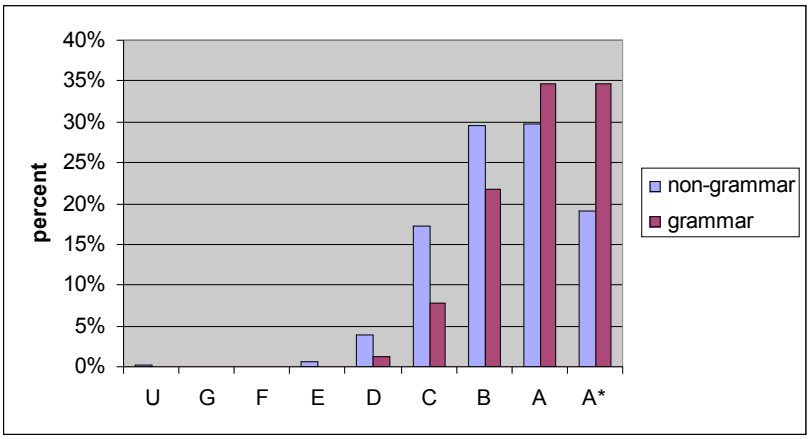
Chemistry

Number of grammar school entries 5,014

(23.7% of the cohort)

Number of non-grammar school entries 12,273

(2.4% of the cohort)



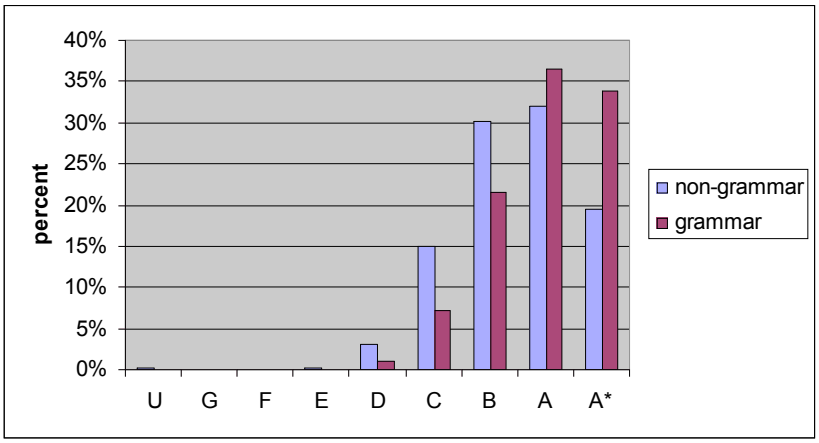
Biology

Number of grammar school entries 5,017

(23.7% of the cohort)

Number of non-grammar school entries 12,347

(2.4% of the cohort)



We see that the entry numbers in both the grammar schools and non-grammar schools were very consistent, suggesting that it is the same cohort of pupils at both types of school who took the three sciences. The percentages given in the boxes in Figure 32 are relative to the whole grammar school and non-grammar school cohorts. Relative to the cohorts with three level 5s or higher at KS2, we would expect these percentages to be higher, and they are at a consistent 40% for grammar schools and 18% for non-grammar schools, but they haven't increased substantially on the percentages shown in Figure 25.

It is apparent from the graphs in Figure 32 that the grammar school pupils continued to perform better than those from the non-grammar schools. In all three sciences about 70% of the grammar school pupils obtained a grade A or A* compared to about 50% in the non-grammar schools. The percentages achieving at grade C or higher were about 99% for the grammar schools and 96% for the non-grammar schools, which is higher than the 85% shown in Figure 25 (p197), which is to be expected.

For completion we compare the performance in single science for those pupils who had three level 5s at KS2. Although the numbers are low, it is surprising that any of these pupils would take the single science qualification 5 years after their KS2 level 5s achievement.

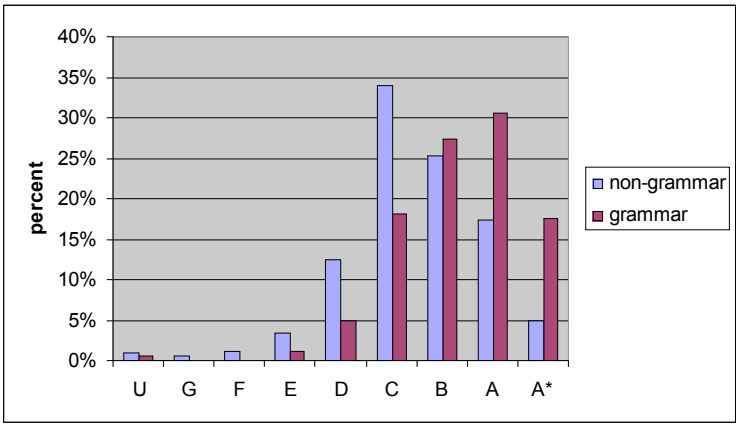
Single Science

Number of grammar school entries 183

(0.9% of the cohort)

Number of non-grammar school entries 2,297

(0.4% of the cohort)



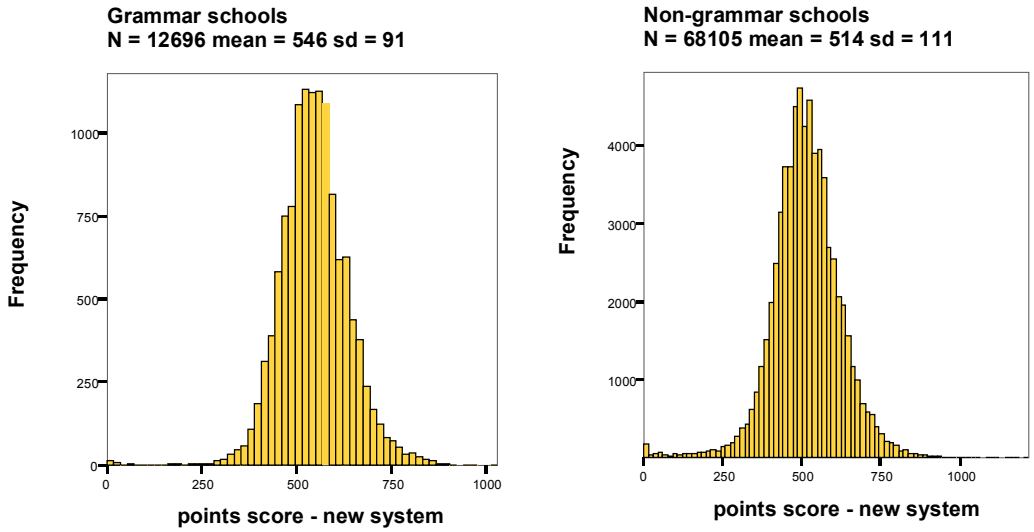
The grammar school pupils are again seen to have out performed the non-grammar school pupils in terms of the percentages of grades achieved, with 48% achieving a grade A or A* compared to 23% respectively. At grade C or higher, the grammar schools achieved a 93% pass rate compared to 81% for the non-grammar schools.

Again we have to question whether we are comparing like with like here. It is possible that although we are comparing students with three levels 5s or higher at KS2 the grammar school students may still be slightly more able than the non-grammar school students at the outset (see Section 4.1.2, p108) and that this could account for the differences in GCSE grades.

8.2.4. Matching by Key Stage 2 Level: Comparing Aggregated Outcomes

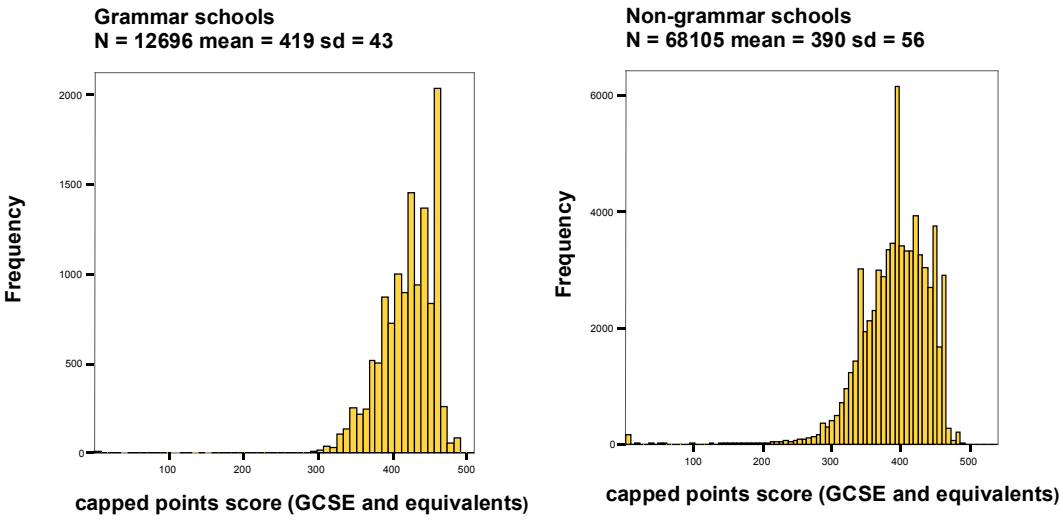
Aggregated Raw GCSE Results for pupils achieving three level 5s at KS2.

Figure 33: DCSF GCSE and equivalents Total Points Score



The mean value for the grammar schools is still higher than for the non-grammar schools although the difference, about 30 points is much smaller, as would be expected. Similarly the standard deviation for the grammar schools has remained smaller but the standard deviation for the non-grammar schools has decreased by about 40 points. Interestingly the mean for the non-grammar schools is about equal to that for all pupils in grammar schools. The long tail of low achievement seen in Figure 26 (p200) for the non-grammar schools has disappeared as might be expected and both distributions are symmetrical about their mean values.

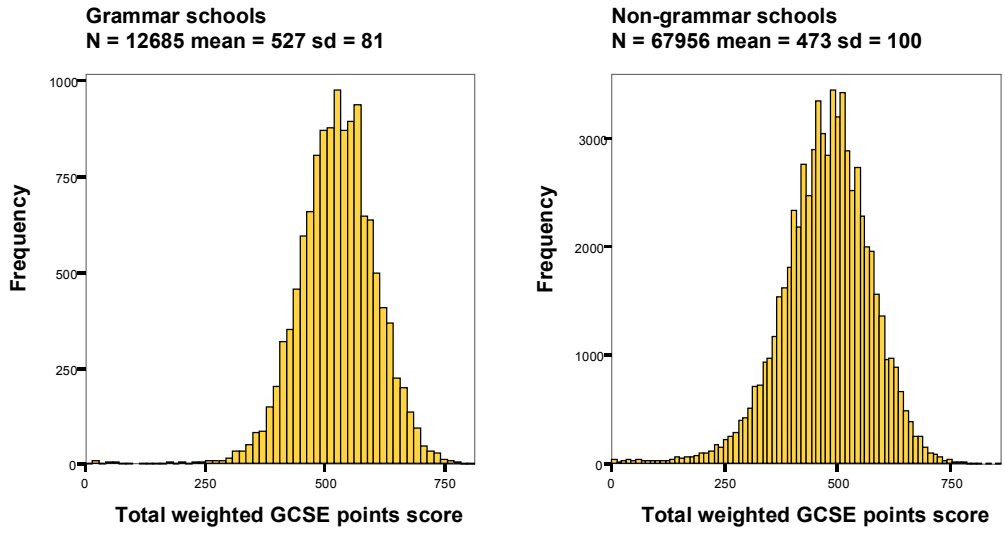
Figure 34: DCSF Capped Points Score



The ceiling effect is now apparent for both types of school but the mean value for the grammar schools is still higher by about 30 points. Although for both types of school there are relatively few pupils with below 300 points it is more noticeable for the non-grammar schools. It is notable for pupils in this category that there are still some who scored zero or below 30 points, there

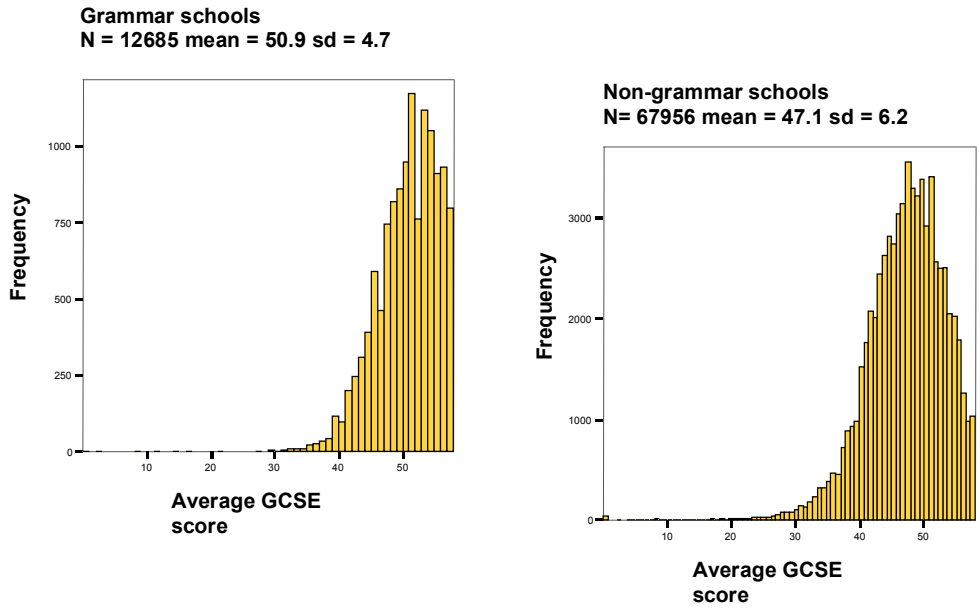
being 30 such pupils from the grammar schools and 218 such pupils from the non-grammar schools. At the high performance end 1785 grammar school pupils scored 464 points or above (14% of this cohort) compared with 3068 non-grammar school pupils (4.5% of this cohort).

Figure 35: Total Weighted GCSE Point Score



The distributions in Figure 35 are very similar to those of Figure 33, which might be expected as the more able pupils as shown by their KS2 achievement are likely at KS4 to have taken GCSE subjects. The mean of the grammar school pupils is higher by 54 points, or between a grade A and A* with the standard deviations differing by 19 points. The distributions are symmetrical about their mean values and relatively few pupils scored below 300 points in both types of school.

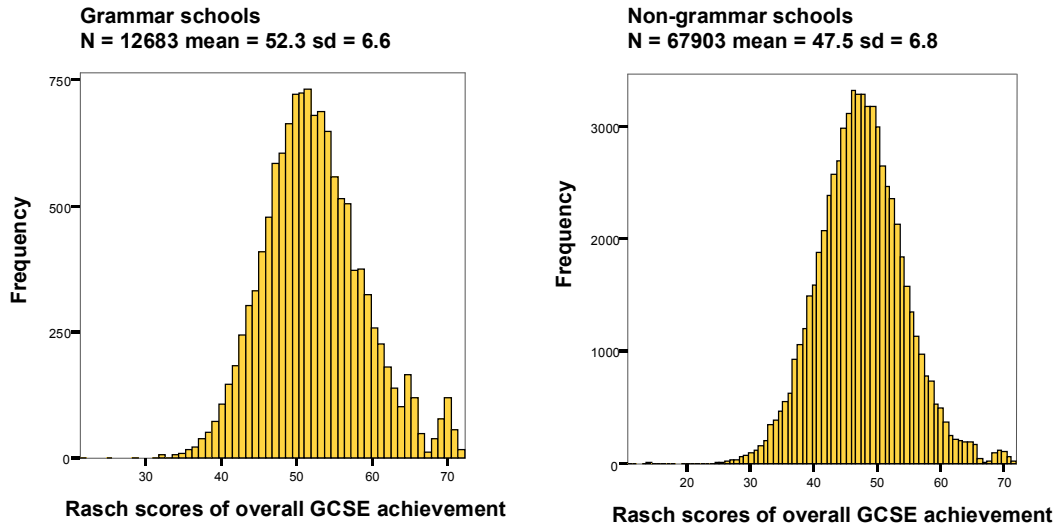
Figure 36: Average Points Score



The average points score for both types of school has increased and more so for the non-grammar schools although it is still smaller. The mean for the

grammar schools at about 51 represents just under a grade A, and that for the non-grammar schools 47, just over a grade B. The standard deviation has also remained smaller, indicating greater consistency of results in the grammar schools. A ceiling effect is now evident in both distributions. For the grammar schools there were 293 at the maximum of 58 with about 47% of this cohort averaging 52 (a grade A) or more points and about 85% averaging 46 (a grade B) or more points. The mean number of GCSE subjects taken was 10.3 with a standard deviation of 1.2. For the non-grammar schools there were 406 pupils at the maximum of 58 with about 24% of this cohort averaging 52 (a grade A) or more points and about 62% averaging 46 (a grade B) or more points. The mean number of GCSE subjects taken was 10.0 with a standard deviation of 6.2. The performance of the grammar school pupils is still seen as relatively better than that of the non-grammar school pupils.

Figure 37: Rasch score of overall GCSE achievement



The two graphs shown in Figure 37 are very similar, and although the mean for the grammar school pupils is still higher than that of the non-grammar school pupils, they are differ by only about five points as opposed to about 13 for the full cohorts of Figure 30; the standard deviations are approximately equal. The two graphs are now similar at the high achievement end of the scale, probably indicating that the most able pupils from both types of school were entered for the more difficult subjects.

8.3. *Value-added approaches*

The detailed results of the various regression and multilevel models we used are given in Chapter 10; *Appendix: Results from all models.*

8.3.1. Results of the regression and multilevel models

In analysing the results from these models, we were looking to see if various combinations of explanatory variables made a significant difference to the observed grammar school effect and its controls, and whether there was a significant difference in results between the two types of model.

Some results from the statistical models are summarized in Table 40 to Table 42, below.

Table 40: Grammar school effect, calculated from different models: Comparing grammar schools with other schools using individual subjects

Outcome	Explanatory variables entered	R^2 (OLS)	Difference between grammar schools and others (GCSE grades)		
			OLS	ML	ML (var slopes)
Mathematics	KS2 average level	0.45	0.86	0.97	1.15
	KS2 average mark	0.49	0.65	0.73	0.80
	Pupil level variables	0.51	0.56	0.63	0.70
	Pupil and school level variables	0.52	0.15	0.17	0.15
English	KS2 average level	0.42	0.75	0.85	0.98
	KS2 average mark	0.42	0.63	0.72	0.77
	Pupil level variables	0.48	0.53	0.62	0.62
	Pupil and school level variables	0.49	0.04	0.05	0.08
Double Science	KS2 average level	0.40	0.74	0.77	0.88
	KS2 average mark	0.43	0.60	0.62	0.67
	Pupil level variables	0.46	0.51	0.53	0.58
	Pupil and school level variables	0.47	0.06	0.05	0.08
History	KS2 average level	0.36	0.62	0.80	0.72
	KS2 average mark	0.39	0.49	0.65	0.58
	Pupil level variables	0.44	0.40	0.55	0.45
	Pupil and school level variables	0.45	-0.18	-0.18	-0.20
French	KS2 average level	0.37	0.80	0.80	0.85
	KS2 average mark	0.39	0.69	0.70	0.72
	Pupil level variables	0.44	0.66	0.67	0.65
	Pupil and school level variables	0.45	0.17	0.30	0.2
English Literature	KS2 average level	0.33	0.61	0.70	0.80
	KS2 average mark	0.34	0.51	0.60	0.65
	Pupil level variables	0.42	0.42	0.50	0.52
	Pupil and school level variables	0.41	-0.11	-0.10	-0.05

The numbers in Table 40 are the coefficients resulting from the regression analysis. These coefficients represent fractions of a GCSE grade. Thus, if we consider the first row of data as an example, we are taking the grade achieved in GCSE mathematics as the outcome. We start with a model in which only one explanatory variable, 'KS2 average level', has been entered, and this accounts for '0.45' (i.e. 45%) of the variation in our outcome. The three right-hand columns of the table show the differences between the achievements of pupils in grammar schools and those in non-grammar schools in mathematics, when the explanatory variable 'KS2 average level' has been taken into account. The three different values are the result of running three different types of statistical model.

These three statistical models differ in the features that they take into account. The 'OLS' model uses ordinary least squares regression to estimate the overall relationship between the explanatory variable(s) and the outcome, effectively plotting a line of best fit between them. In fact, we ask it to plot two parallel lines, one to represent the relationship for pupils in grammar schools, the other for those in non-grammar schools. The distance between these two lines gives us the 'Difference between grammar schools and

others', which is given as '0.86'. This may be interpreted as meaning that, when we take account of 'KS2 average level' in an 'OLS' model, pupils in grammar schools achieve on average 0.86 of a grade better than those in non-grammar schools in GCSE mathematics.

The 'ML' model uses multilevel modelling to estimate the relationship. Essentially the difference is that this model takes account of the fact that pupils are grouped within schools and so estimates the relationship within each school. Effectively, it plots a line of best fit for each school, though in this model the lines are all constrained to be parallel. It then compares the average heights of the lines for grammar and non-grammar schools. The interpretation is essentially the same as for the previous model: when we take account of 'KS2 average level' in an 'ML' model, pupils in grammar schools achieve on average 0.97 of a grade better than those in non-grammar schools in GCSE mathematics.

The final model, 'ML (var slopes)' is another multilevel model, but now the lines of best fit for each school are not constrained to be parallel – their slopes are allowed to vary. Comparing the performance of grammar and non-grammar school pupils in this model is more complex, since the size of the difference can vary according to the characteristics of the pupil. The figures shown in the table (e.g. '1.15' in the first row) represent the difference estimated for an 'average' pupil. The interpretation is again the same: according to this model, pupils in grammar schools achieve 1.15 grades better than those in non-grammar schools.

Table 41: Grammar school effect, calculated from different models: Comparing grammar schools with other schools using average GCSE weighted points and GCSE points with difficulty accounted for using Rasch technique.

Outcome	Explanatory variables entered	R^2 (OLS)	Difference between grammar schools and others (GCSE grades)		
			OLS	ML	ML (var slopes)
Average GCSE weighted points	KS2 average level	0.46	0.87	0.98	1.13
	KS2 average mark	0.48	0.70	0.80	0.85
	Pupil level variables	0.54	0.59	0.68	0.70
	Pupil and school level variables	0.55	0.05	0.07	0.03
GCSE points (using Rasch technique for subject difficulty)	KS2 average level	0.51	1.03	1.13	1.18
	KS2 average mark	0.54	0.87	0.97	0.80
	Pupil level variables	0.59	0.78	0.87	0.72
	Pupil and school level variables	0.60	0.28	0.30	0.27

Table 42: Grammar school effect, calculated from different models: Comparing grammar schools with other schools using DCSF total points and capped total points, and total points calculated from weighted GCSE qualifications.

Outcome	Explanatory variables entered	R^2 (OLS)	Difference between grammar schools and others (GCSE grades)		
			OLS	ML	ML (var slopes)

Total points DCSF system	KS2 average level	0.37	7.94	9.20	11.67
	KS2 average mark	0.38	5.52	6.63	7.00
	Pupil level variables	0.42	4.25	4.72	5.20
	Pupil and school level variables	0.42	-0.25	0.28	0.86
Capped total points DCSF system	KS2 average level	0.40	6.38	7.27	7.77
	KS2 average mark	0.41	4.91	5.75	5.60
	Pupil level variables	0.46	3.92	4.42	4.17
	Pupil and school level variables	0.47	-0.11	0.00	-0.15
Total GCSE weighted points	KS2 average level	0.43	12.29	14.17	18.90
	KS2 average mark	0.45	10.05	11.87	13.37
	Pupil level variables	0.51	8.55	10.25	11.78
	Pupil and school level variables	0.52	0.51	0.23	3.27

The coefficients in Table 41 represent a fraction of a GCSE grade averaged across all subjects taken. In Table 42 the coefficients have been presented in total points form (ie six points to one GCSE grade). The results in Table 40 to Table 42 show that there is a great variation in the grammar school effect depending on the explanatory variables and model used (see Chapter 9 for a discussion on which model/variables we should make use of). The results range from around plus one GCSE grade to around zero and some small negative results in some cases when pupil and school level variables are input to the models. i.e. at best, grammar schools appear to be improving each GCSE result by a grade, but as we add in more explanatory variables this effect gradually disappears towards zero. Looking at Table 40 we can see that the grammar school effect is at its largest with French and Mathematics, and lowest with History and English Literature. These results should also be considered from the perspective of the subject entries; these are shown in Table 43. The total grammar school cohort is 21,122 pupils and the total non-grammar school cohort is 512,761 pupils.

Table 43: subject entries in 2006 for grammar schools and non-grammar schools

	Grammar Schools		Non-grammar Schools	
Mathematics	21033	99.6%	498490	97.2%
English	21062	97.2%	494894	96.5%
Double Science	13679	64.8%	327881	72.7%
History	11765	55.7%	161739	31.5%
French	12456	59.0%	156881	30.6%
English Literature	20823	98.6%	430921	84.0%

It is notable from Table 40 and Table 43 that although a much higher proportion of pupils from grammar schools were entered for History and French, the grammar school effect is much more noticeable for French.

The results in Table 40 indicate that in general the grammar school effect is higher when subject difficulty is taken into account in the average points score based on GCSE subjects only. It is likely that grammar school pupils tend to take the more difficult subjects as found from the Rasch technique although further investigation would be necessary to substantiate this. Table 42 shows differing results for total points and capped points using the DCSF

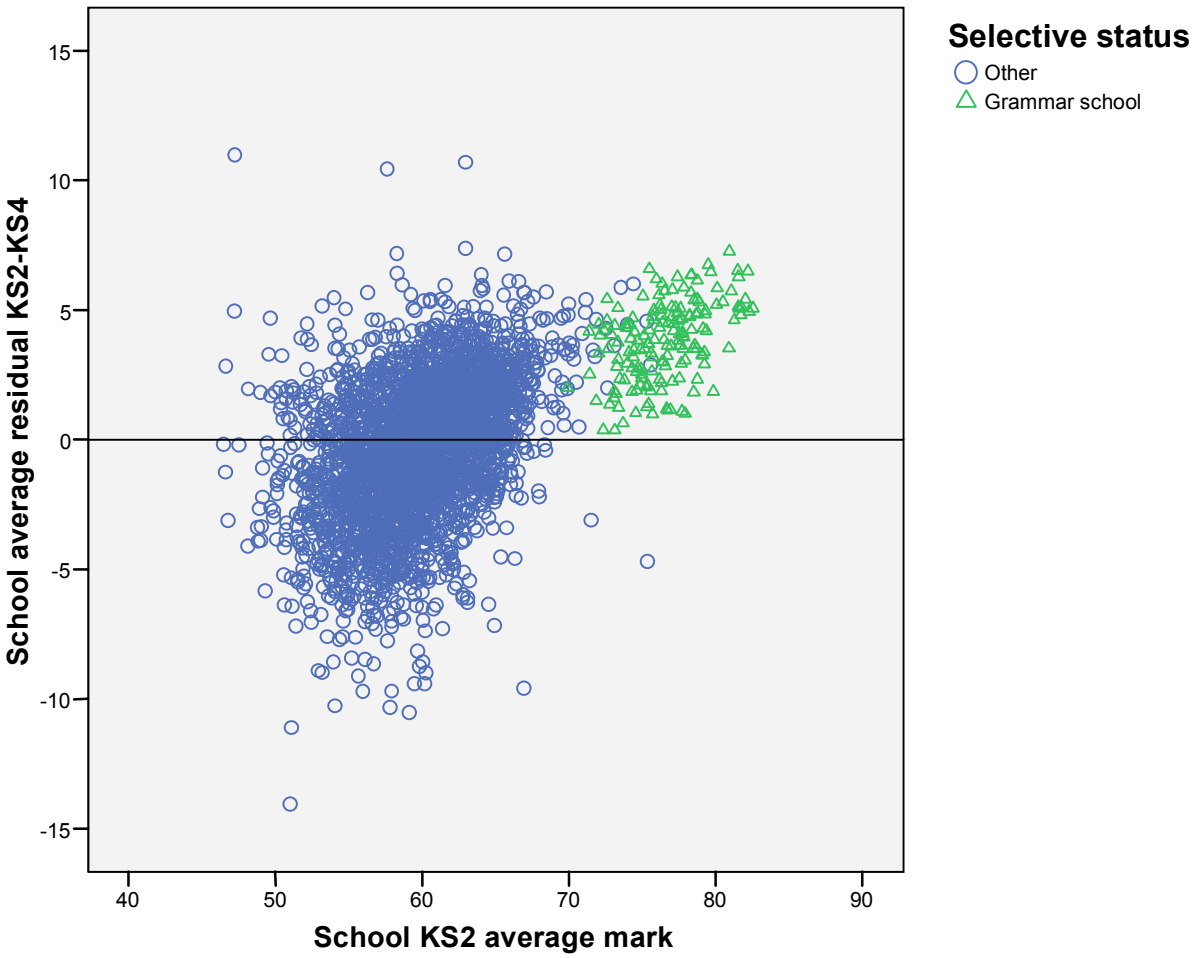
data, with the capped points showing a reduction in effect. It is likely that this reduction is due to the ceiling effect illustrated in Figure 27. The DCSF points are calculated from GCSEs and their equivalents, whereas the total GCSE weighted points have been calculated for GCSE subjects only. The grammar school effect is notably higher, and this is probably due to the likelihood that all grammar school pupils take GCSE subjects. Again this would warrant further investigation.

8.3.2. Modelling compositional effects

The issue of whether school level compositional variables should be included in the model has been discussed in 8.1.2 (p188).

Figure 38 shows the relationship between a school's average KS2 mark and its average residual in a model with a pupil's average KS4 points as the outcome and the pupil's KS2 average mark entered as the only explanatory variable.

Figure 38: School composition and residuals from pupil-level regression



It is clear that there is an overall tendency for schools with higher average KS2 marks to have higher residuals – the compositional effect. However, it is also clear that the two kinds of school, grammar schools and others, are two quite separate groups. Figure 38 indicates that all the grammar schools on

average had positive residuals, and in general their average KS2 mark is higher. The performance of the grammar schools is also much more consistent than that of the other schools.

There are three possible ways one could incorporate a compositional effect into an overall estimate of the grammar school effect. One would be to treat all schools as members of a single group and calculate an overall compositional effect. A second would be to calculate separate compositional effects for each type. A third would be to calculate the compositional effect for non-grammar schools and extrapolate to grammars. None of these solutions seems very satisfactory, and the answers they give are very varied. The estimates of the grammar school effect from some OLS (ordinary least squares) models using each method are shown in Table 44, together with the estimate from a model with no compositional variables included (*Model 0*).

Table 44: Effects of different ways of modelling the 'compositional effect'

	Model 0	Model 1	Model 2	Model 3	Model 4
R-squared	0.50	0.52	0.52	0.49	
Intercept	35.57	19.74	19.61	19.82	19.35
ave KS2 score	7.23	6.80	6.80	6.86	6.70
school ave KS2 score		0.28			
school ave KS2 score (grammars)			0.07		0.25
school ave KS2 score (non-gram)			0.28	0.28	0.28
grammar school	3.49	-0.79	14.82	-0.84	2.53
grammar school (GCSE grades)	0.58	-0.13	2.47	-0.14	0.42

Model 0: No compositional variables
Model 1: All schools treated as one group
Model 2: Separate compositional effect for grammar and non-grammar
Model 3: Restricted to non-grammar, then extrapolated to grammar
Model 4: As Model 2, but multilevel with variable school KS2 slopes

Outcome variable = average points

Model 1, which is perhaps most similar to what has conventionally been done (e.g. Schagen and Schagen, 2005) provides a negative estimate of the grammar school effect, as does *Model 3*. However, given the picture in Figure 38, it is hard to see how either of these models could really be defended; grammar schools and other schools are so different that to treat them as the same or to extrapolate from one group to the other seems hard to justify. *Model 0* does not so much solve this problem as ignore it, leaving compositional effects out of the model altogether.

Model 2 has a certain appeal, as it directly models the relationship for each group of schools separately. However, the effect of including a compositional effect for non-grammar schools is to reduce the slope of the individual KS2 variable (from 7.23 in Model 0 to 6.80) for all schools, including the grammar schools whose compositional effect is almost zero and whose pupils are all

towards the top end of the KS2 scale. As a result, the residuals of the latter are increased quite substantially, giving an estimate of the grammar school advantage of some two and a half grades per GCSE subject taken. This is equivalent to saying that a pupil who might be expected to achieve all C grades in a comprehensive school would get As and A*s if instead they attended a grammar school. Although it is dangerous to judge the adequacy of a model by its results, such a big difference does seem hard to believe.

In response to this implausibly large apparent effect, we then fitted a further model (*Model 4*). This is a multilevel model with school KS2 slopes allowed to vary. Although the fixed coefficient of KS2 is even smaller than in the other models (6.70), the variation in slopes for individual schools allows the overall grammar school effect to shrink back to just under half a grade. Interestingly, in this model the size of the compositional effect for grammar schools is now about the same as that for other schools.

In the analyses presented in this chapter we have used *Model 0* and *Model 1* (including their equivalents in multilevel models) for comparative purposes, despite these reservations. This is partly because to do so duplicates previous approaches and follows established conventions of the field (e.g. Schagen and Schagen, 2005), but also because *Model 1* seems to provide a lower bound or 'worst case scenario' for grammar schools.

8.3.3. Correction for reliability

A variation on the analysis is to make a correction for the reliability of the independent variable. It has long been known that in comparing outcomes for two groups that are initially quite different, a bias occurs if the reliability of the pre-test measure is less than 1. We discussed this issue in Chapter 4.

For ease of comparison we use the OLS model with average GCSE points as the outcome and just one explanatory variable, KS2 score, to illustrate (see Table 41, second row of data for the OLS model). When no adjustment is made for reliability, the grammar school effect in this model is 0.70 of a GCSE grade: pupils in grammar schools achieve on average 0.70 of a grade per subject better than pupils in non-grammar schools who started with the same KS2 score.

However, as explained in Section 4.1.2 (p108), even for two pupils with the same recorded KS2 scores, knowing that one went on to a grammar school while another did not means that the former is likely actually to be more able. In other words, we are not strictly comparing like with like.

If we assume a reliability of 0.9 and adjust pre-test scores towards their respective group means before running the regression model, the grammar school effect shrinks correspondingly to 0.55 of a grade. With reliability 0.7 this becomes 0.12, suggesting very little difference between the two groups.

8.3.4. Different comparison groups

So far, all the models we have considered have been applied to comparisons between the performance of pupils in grammar schools and those in other schools. We now turn to the other ways that we might choose to divide pupils as listed in Section 8.1.4 (p195).

Table 45: Estimates of the differences in the performances of different comparison groups

Outcome = GCSE points (using Rasch technique for subject difficulty)	Estimates of difference in performance (from OLS models)	
Comparison groups Based on 533883 pupils; year 11 2005/06 3115 schools	Model with all pupil-level variables	Model with all pupil- and school-level variables
1. Grammar schools vs all other schools Schools 164 vs 2951 Pupils 4.0% vs 96.0%	0.78	0.28
2. Grammar schools and >20% creamed vs all others Schools 325 vs 2790 Pupils 8.4% vs 91.6%	0.33	0.09
3. Grammar schools and >5% creamed vs all others Schools 709 vs 2406 Pupils 20.8% vs 79.2%	0.15	0.04
4. Grammar schools and >1% creamed vs all others Schools 1353 vs 1762 Pupils 41.4% vs 58.6%	0.10	0.03
5. Selective LAs (i.e. at least 10% in grammar schools) vs others Schools 561 vs 2554 Pupils 12.4% vs 87.6%	0.12	0.02
6.Grammar schools and secondary moderns vs comprehensives Schools 343 vs 2772 Pupils 8.4% vs 91.6%	0.32	0.09

Table 45 summarizes the results for the various ways we have chosen to define 'selective systems', as discussed in Section 8.1.4. The table also shows the number of state schools and proportions of pupils in each category. The choice of level of creaming at greater than 20%, greater than 5% and greater than 1% is arbitrary, but these were chosen to investigate the 'selective system' effect in three possible definitions of 'selective system'. The 'selective system' effect on the grammar schools is measured in GCSE grades. The first result essentially shows that grammar schools are having a positive effect of about three quarters of a GCSE grade when pupil level variables are included and this reduces to about one quarter of a grade when pupil and school level variables are both included.

Results 2 to 4 allow us to look at whether the grammar schools are having a detrimental effect on the schools that they cream from, using the definition of creaming detailed in Section 7.1 (p153). If the grammar schools were having a detrimental effect we would expect to see some negative numbers in the model with all pupil level variables. We see this is not the case and the effect remains positive for all our definitions of 'selective systems' meaning overall students are doing better in 'selective systems' than 'non-selective systems'.

However, we do see the effect of the grammar schools decrease to about one third of a grade when schools that are highly creamed (>20%) are added to

them for comparison with all other schools. At this level of creaming we might expect a result similar to comparing grammar schools and secondary modern schools together with all other schools. We see in result 6 that the effect is indeed very similar. Results 3 and 4 show that as more schools are included with the grammar schools the grammar school effect is reduced. This is due to dilution (i.e. due to more schools being included, with progressively fewer creamed pupils, limiting the difference between the two comparison groups). Grammar schools together with schools that are creamed at a level of over 1%, when compared with all other schools, shows just a small positive effect.

Result 5 compares all schools in selective LAs with all other schools. Although we have criticised the LA as an inappropriate unit of analysis, it is included here for completion. The choice of 10% is arbitrary but consistent with what other researchers have chosen. We see the grammar school effect is again positive, but small.

Thus we were unable to find a grouping of schools that showed the grammar schools as having a negative effect. This is not to say that there is no such effect, but any negative effect on other schools may not be large enough to make the overall effect in the regression analysis become negative.

When school-level variable data is added there is essentially no difference between the comparison groups 2 to 5; the grammar school effect stays around zero for these five definitions of selective systems. The only discernable difference, of about one quarter of a grade, is in comparison group 1 where grammar schools alone are compared to all other schools. This may be due to subject difficulty; it is likely that relative to the size of the year group, much larger cohorts of pupils in a grammar school take the more difficult subjects compared to other schools. This would warrant further investigation. Here, the effect of creaming is considered further in Figure 39.

Figure 39: Variation of schools' value-added performance with level of creaming

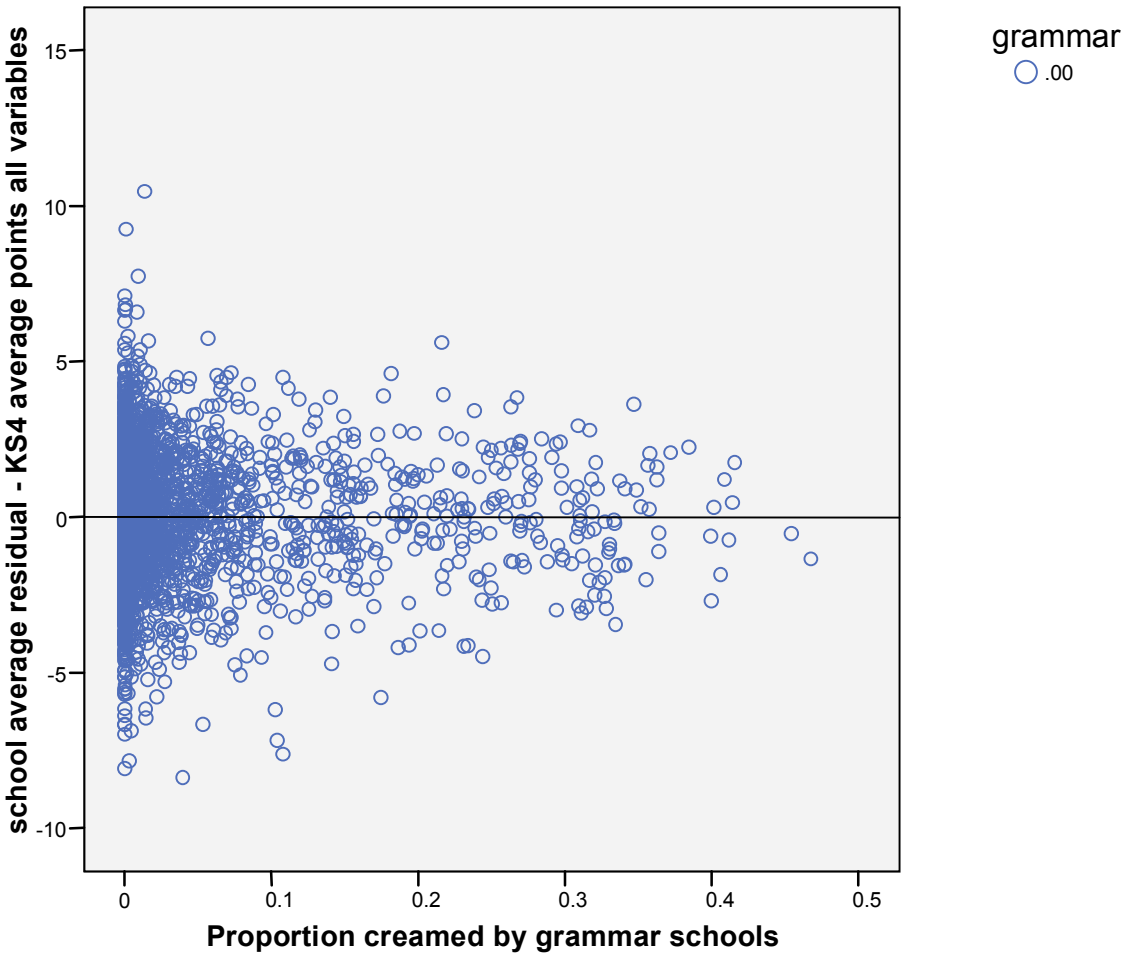


Figure 39 shows that there is really no relationship between the amount of creaming experienced by a school and its performance, in value-added terms. The value added as measured on the vertical axis in Figure 39 is the residual from regression of a pupil's GCSE average points score on the pupil's average KS2 mark averaged for each school. In fact the correlation between this measure and the proportion creamed by grammar schools is 0.003. In the light of this failure to find any association between performance and creaming, it is perhaps not surprising that comparisons 2 to 6 in Table 45 show a similar picture.

8.3.5. Interaction effects: FSM and Grammar Schools

This section presents an investigation of the interaction between grammar school attendance and Free School Meals (FSM) status. We have seen fairly consistently in the statistical models that being eligible for FSM is associated with just under half a GCSE grade poorer performance than non-FSM pupils with the same prior attainment and other characteristics. It seems appropriate to ask whether this disadvantage is more or less in grammar schools than in other schools. This question can be addressed by including an 'interaction term' in the model.

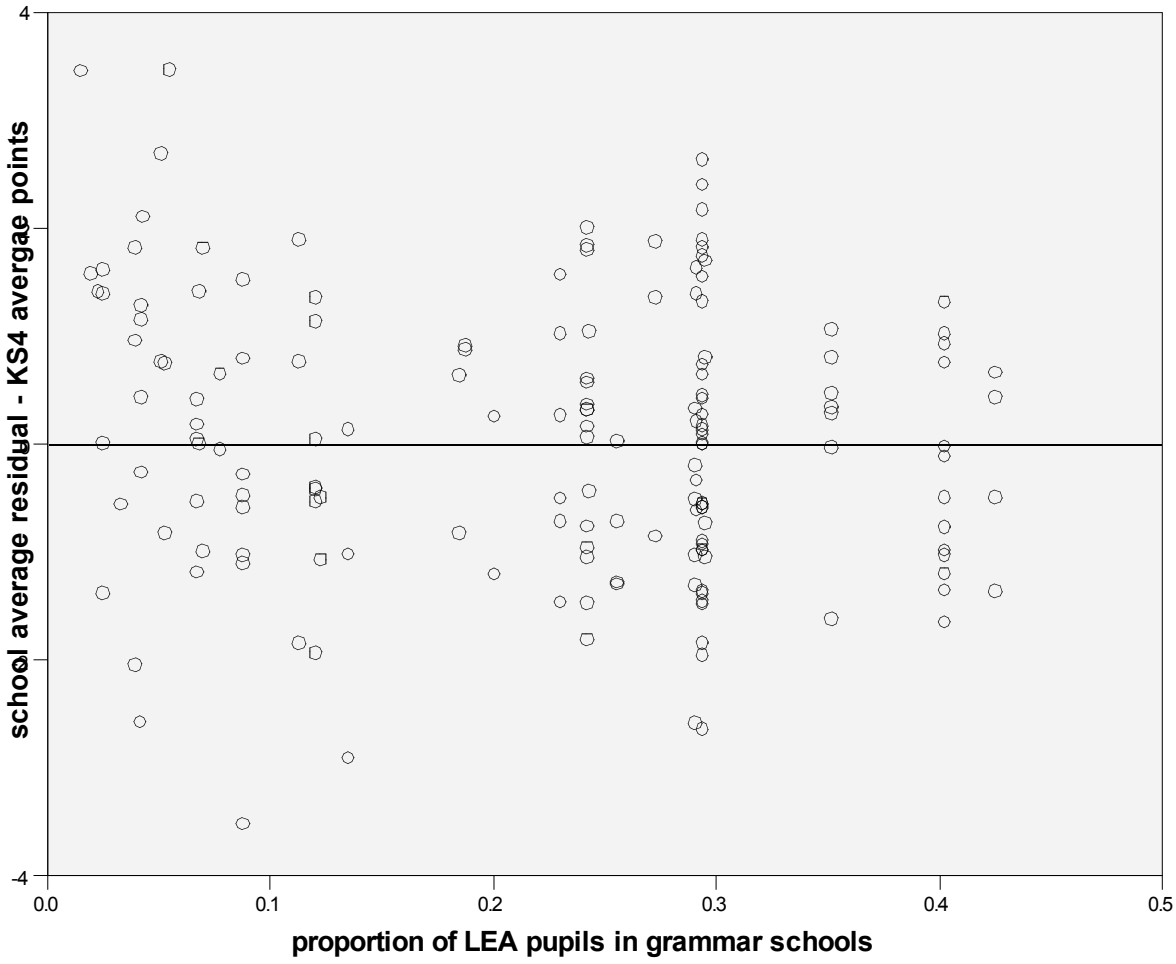
When we do this, we find that pupils eligible for FSM appear to suffer marginally less educational disadvantage if they attend grammar schools. The difference is equivalent to about one-eighth of a GCSE grade; although this is statistically significant, it is certainly not large. It also seems possible that FSM pupils in grammar schools may typically be quite different from FSM pupils as a whole in ways that are not well measured, so we should be cautious about interpreting this as a strong endorsement of grammar schools.

8.3.6. Effectiveness of grammar schools in LAs with different levels of selectivity

A further question we might ask is whether the overall effectiveness of grammar schools is explained by the fact that some types of grammar school are particularly effective. One variable that might be significant here is the level of selectivity in the LA as a whole.

To answer this we calculated the proportion of secondary pupils in each LA where there is at least one grammar school and compared it with the mean residuals for the grammar schools. The residuals are the mean residuals for the school from an OLS model with the individual pupils average GCSE KS4 points score in regression with each pupils KS2 score. These residuals are a measure of the overall effectiveness of the school (value-added). The result is shown in Figure 40. It is clear from this that there is no particular relationship; grammar schools are equally likely to be effective, regardless of the level of selection in the LA in which they are located.

Figure 40: Relationship between level of selection and effectiveness for grammar schools



8.4. *Progress made between KS1 and KS2 of grammar school pupils*

Manning and Pischke (2006) present an interesting analysis of data from the 1958 birth cohort National Child Development Study in which they compare the progress made between ages 7-11 of those pupils who then went on to grammar/secondary modern schools with those in comprehensives. They found that those in selective systems were already making more progress, even before the selection had been applied to them. This study was reviewed in section 3.3.4, p93.

Manning and Pischke interpret this result as suggesting that what appears to be a 'selective schooling' effect may actually be a result of some unmeasured differences between pupils in the two types of system. It seems unlikely that a pupil's progress through primary school would be affected by whether they subsequently entered a selective or comprehensive secondary schooling system, especially after controlling for commonly used background variables such as prior achievement and socio-economic factors. However, they did indeed find such a 'selective system' effect on progress from age seven to eleven, after controlling for background variables. The most likely explanation for this is that the model is underspecified – some crucial

variables have been omitted that, had they been included, would have accounted for the different rates of progress. If the model used in primary schools is underspecified, how can we be confident the model used when studying secondary schools is not similarly underspecified?

We have replicated their findings using the national pupil datasets for pupils who sat their GCSEs in 2007 to assess whether our models may be similarly affected by under-specification. This is the first year for which data from Key Stage 1 (taken when these pupils were seven, in 1998) can be matched to their GCSE results.

Table 46: Value-added results measuring the 'grammar school' effect and 'selective system' effects from KS1 to KS2 and KS2 to KS4 using OLS regression

	Grammar schools vs others		Selective LAs vs others	
	R^2	Effect Size	R^2	Effect Size
Primary school: Progress from KS1 to KS2				
Model1: KS1 avg mark	0.60	0.37	0.58	0.04
Model 2: all pupil vars	0.60	0.34	0.59	0.03
Model 3: all pupil and sch level vars	0.61	0.34	0.60	0.02
Secondary school: Progress from KS2 to KS4				
Model1: KS2 avg mark	0.53	0.37	0.52	0.06
Model 2: all pupil vars	0.58	0.29	0.58	0.05
Model 3: all pupil and sch level vars	0.59	-0.01	0.59	-0.01

Our results are based on the same models used for the 2006 dataset analysis. In the KS1 to KS2 analysis the outcome is the pupil's average mark across maths, English and science. Model 1 controls for average KS1 marks, model 2 controls for all pupil level variables (KS1 marks, plus sex, FSM status, IDACI deprivation index, and ethnicity) and model 3 controls for all pupil and school level variables (as model 2, plus school averages of KS1, FSM and IDACI). In the KS2 to KS4 analysis the outcome is the pupil's average GCSE grade. Model 1 controls for KS2 marks, model 2 controls for all pupil variables, model 3 controls for all pupil and school level variables. The analysis for each phase of schooling was repeated twice, measuring two different effects. We firstly measured the "grammar school effect" to mirror the main analysis in this report. Secondly we measured the "selective system effect" to mirror Manning and Pischke's results. We define selective systems to be LAs with 10% or more students attending grammar schools. The results are summarised in Table 46. Full results can be found in Appendix 10.3.

As the outcomes at KS2 and KS4 are measured in different units, we have converted both to an 'effect size' statistic: the difference between groups divided by the population standard deviation (Coe, 2002). The results show that the 'grammar school' effect and the 'selective system' effect are almost identical in all bar one of the models. Only for the comparison between pupils in grammar schools and other schools in the full model (3) with all pupil and

school level variables included is there a difference in the effect size. Given the problematic nature of 'compositional effects', as discussed in section 8.3.2 above, which are likely to behave differently in primary and secondary schools, this difference may not be too surprising.

Overall, therefore, we appear to have largely confirmed the findings of Manning and Pischke (2006) that the apparent advantage conferred by grammar school attendance on GCSE performance can be seen to be mirrored exactly by a comparable advantage in achievement at the end of primary school, before those pupils even set foot in a grammar school. It is possible, of course, that the knowledge that they are about to attend a grammar school causes pupils in Year 6 to put in an extra spurt and improve their Key Stage 2 performance beyond that of their peers. However, it seems much more plausible that some unmeasured (or inadequately measured) characteristic of those who succeed in gaining entry to grammar school, such as greater social or cultural resources, greater academic ability or enhanced motivation, is responsible for their tendency to make greater progress during both primary and secondary school. Hence, what appeared in our earlier analyses to be a 'grammar school' effect now looks more like a selection effect, masked by inadequately specified models. Grammar schools have selected those pupils who would be expected to make more progress, rather than caused their pupils to perform better than they would have done. The pupils who went to grammar schools were already making better progress in their primary schools so it seems likely that they would have continued to make better progress during secondary school whether or not they had gone to a grammar school.

8.5. Propensity Score Matching

The final analysis presented in this chapter uses the method of Propensity Score Matching (PSM). This approach attempts to identify matched sub-samples from the two groups (grammar and non-grammar school pupils) in order to be able to make a direct comparison of their outcomes. The matched sub-samples are selected on the basis of their 'propensities' to attend a grammar school. In other words, if we can find variables that reliably predict whether a particular individual is likely to attend a grammar school or not then we will have taken account of the initial differences that existed between pupils in the two groups. We can then choose a sub-sample from each group in such a way that individuals are matched on these predictor variables.

8.5.1. Creating the matched sub-samples

The obvious variable to predict whether a pupil is likely to attend a grammar school is their ability. Of course, we have no direct measure of ability, but achievement at KS2 would be expected to be a good proxy for ability. Unfortunately, KS2 scores alone are a poor predictor of grammar school attendance since the vast majority of even those with high KS2 scores do not attend grammar schools.

The variable that significantly improves the prediction is the ward level percentage of pupils attending grammar schools. Those pupils who have high KS2 scores and live in wards from which a high proportion of pupils attend a grammar school are significantly more likely to attend them. Inclusion of an

interaction term here (capturing the additional effect of having both high KS2 and high percentage grammar school attendance, over and above the effect of either independently) is important. To allow for the possibility of non-linear relationships between propensity and this interaction term, a further squared and cubed interaction terms were also included.

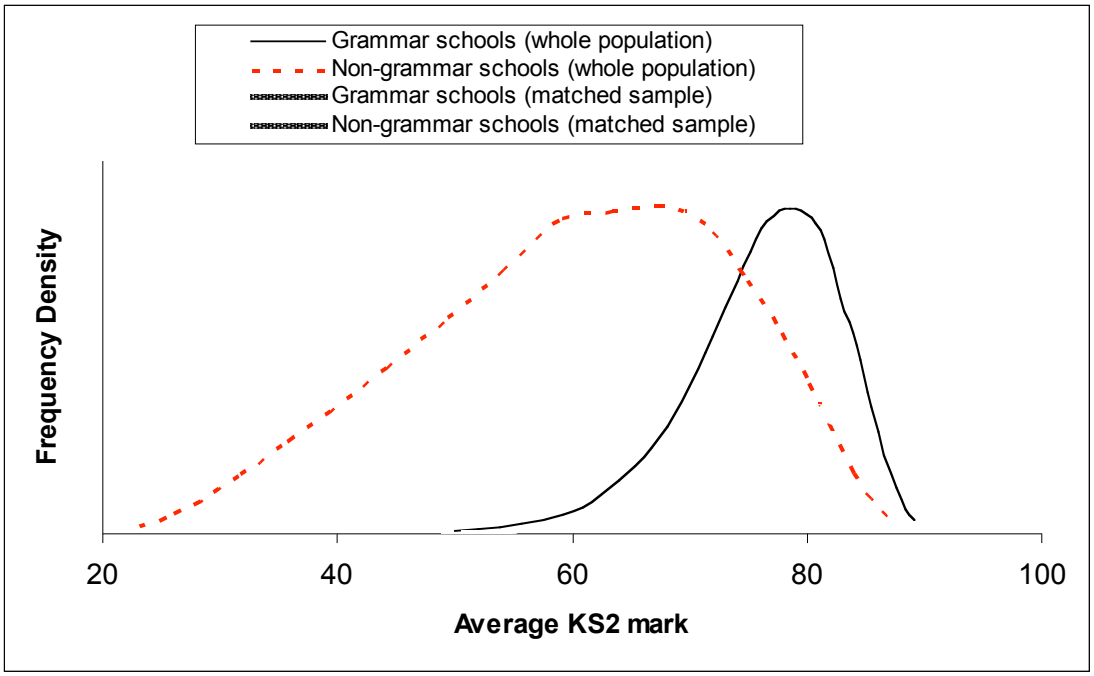
Additional variables included were Free School Meals status, IDACI deprivation index and two ethnicity markers for being 'White British' or 'Indian' which were found to be related to the chances of a pupil attending a grammar school.

Calliper matching was used to obtain the propensity-matched sub-samples. Grammar school pupils were included in the sub-sample if suitable non-grammar school matches could be found, one with propensity less than 0.05 above and one with propensity less 0.05 below. Each of these pupils in the support (non-grammar school) group was weighted 0.5 in all analyses in order to keep the overall sizes balanced. Support pupils were selected with replacement, so it was possible for the same non-grammar school pupil to act as control for more than one grammar school pupil. In this case their weighting was adjusted appropriately.

8.5.2. Comparing the matched sub-samples

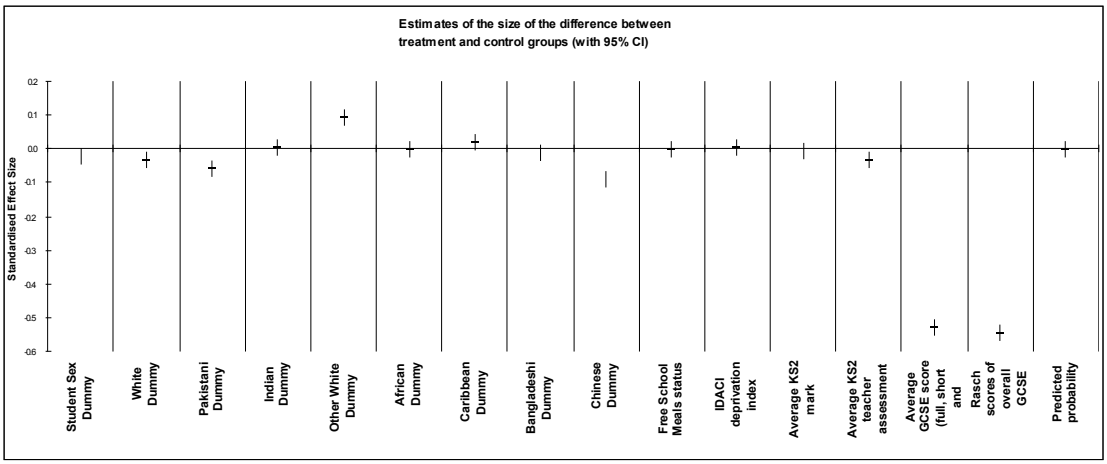
The distributions of KS2 marks for the selected matched sub-samples are shown in Figure 41. It can be seen that for the population as a whole, KS2 marks of non-grammar school pupils are on average somewhat lower than for those at grammar schools, though there is a substantial area of overlap. It is precisely this area of overlap from which the propensity-matched sub-samples have been drawn. Moreover, the process of propensity matching on a range of variables has resulted in two sub-samples that are almost perfectly matched on this particular one. The matched group consists of 14,115 pupils in each group. Comparison of the two sub-samples on all the other variables included in the calculation of the propensity score shows a similar match (see Appendix, p262).

Figure 41: Distributions of KS2 marks for whole population and propensity-matched sub-samples of grammar and non-grammar school pupils



Effect sizes on a range of variables for a comparison between grammar and non-grammar school pupils in the matched sub-sample are shown graphically in Figure 42. For most of the available variables, effect sizes are close to zero, confirming a good match.

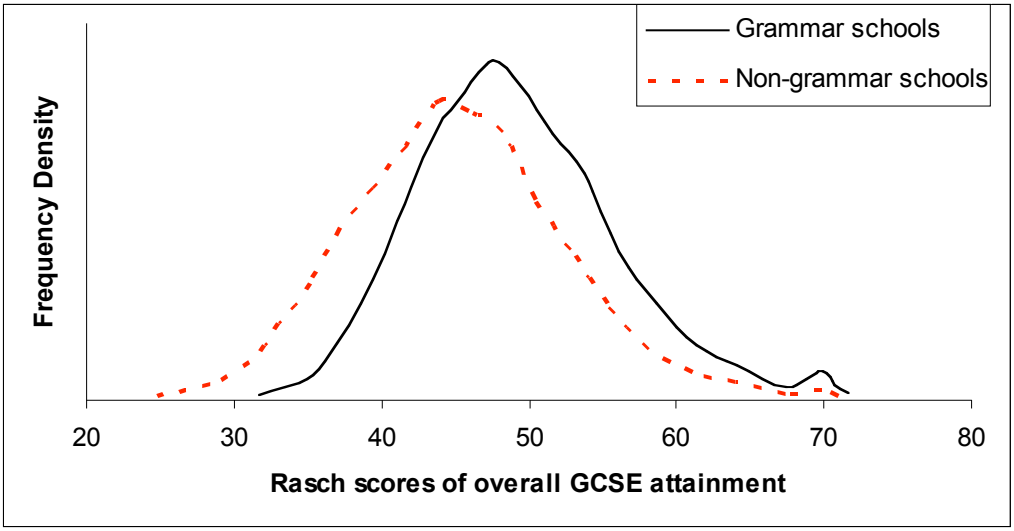
Figure 42: Effect size estimates (with 95% CI) for differences between grammar and non-grammar school pupils in the matched sub-samples



On the two outcome variables, however, the difference is just over 0.5 of a standard deviation, in favour of the grammar school group. These two are the average GCSE points achieved (including short, full and vocational GCSEs) and the Rasch score based on all GCSEs taken (this is an overall measure of achievement which adjusts for the different difficulties of subjects). In terms of GCSE grades, these differences correspond to 0.55 and

0.63 of a grade respectively better performance by grammar school pupils. The distribution of Rasch scores of the two groups are illustrated in Figure 43

Figure 43: Distributions of Rasch achievement scores (QCA points equivalent) of grammar and non-grammar school pupils in the matched sub-sample



Thus although the two groups appear to be well matched on all the baseline variables available, and from everything we know of them should be equally likely to have attended grammar schools, their achievement at GCSE differs appreciably. Those who did actually go to a grammar school have achieved more than half a grade on average better at GCSE.

8.5.3. Regression analysis on the matched sub-samples

One further analysis we can conduct is to adjust again for the small differences that remain between the matched sub-samples, using regression. With one exception, including all relevant explanatory variables in the regression model makes little difference to the overall result. Thus the model with explanatory variables KS2 marks, teacher assessment, FSM, IDACI, sex and ethnicity dummies accounts for 42% of the variance in Rasch scores and estimates the grammar school effect at 0.63 of a grade, exactly as before. Including in addition school-level factors of school average FSM and IDACI and school sex mix change this little (R^2 = 0.42, grammar school effect = 0.51 grade). However, if we include the school average KS2 level, the effect switches and the grammar school effect becomes negative (R^2 = 0.44, grammar school effect = -0.34 grade).

This change may be accounted for by considering the relationship between school average KS2 level and individual achievement, as shown in Figure 44. Within each group there is a tendency for achievement to be higher in schools with higher average KS2 scores. However, the relationship seen on non-grammar schools appears to start again for grammar schools, rather than continuing on the same line. Hence for a given school average KS2 level, the typical achievement is lower in the latter group. The whole issue of whether

and how school composition effect should be applied to two groups that differ so much has been discussed in Section 8.3.2 (p213) and also applies to this analysis.

Figure 44: Scatterplot of Rasch scores with school KS2 average for grammar and non-grammar school pupils in matched sub-sample

PART IV
OVERVIEW

9. Discussion

The first part of this chapter highlights and discusses the key findings from Chapters 6 and 7. We then discuss the different models used in Chapter 8 and attempt to provide some kind of overview of our rather complex findings. Next we relate our results to those in the existing literature and discuss implications for policy and further research.

9.1. Evidence about selection and selective systems

9.1.1. Comparability of selective and non-selective schools

In terms of school-level characteristics, grammar schools are very different from other schools. Without exception, all grammar schools have sixth forms, compared with about half of non-grammar schools. Fewer than 10% of non-grammars are single-sex schools, compared to three-quarters of grammars. Grammar schools also contain higher proportions of specialist schools and Foundation schools.

All these differences make comparisons quite problematic. Any apparent 'grammar school' effect could be, in part at least, a 'single-sex' effect or a 'schools with sixth form' effect, or some combined effect that is greater than the sum of its parts. In theory, one can test for these effects, though if all possible interaction effects are included the number of comparisons quickly becomes huge. However, it is also possible that there are other systematic differences about which we have no data. Comparisons between groups that differ in so many ways should always make us somewhat cautious.

Of course the most obvious difference between grammar and other schools is in the ability levels of their pupils. Although there are plenty of able pupils in non-grammar schools, and hence substantial overlap at the pupil level, in terms of school-level averages, there is no overlap. Grammar and non-grammar schools are like two distinct populations. This raises clear problems for any analysis using school-level variables, since the vastly larger number of non-grammar schools will dominate the model. Combining the two kinds of schools amounts to taking relationships that fit the larger group and extrapolating them to the smaller, so going well beyond what is really justified. The debate over whether (and how) to include school composition variables in the model (see Sections 8.1.2 (p188) and 8.3.2 (p213)) reflects this difficulty.

9.1.2. LAs and the spread of selection

In analysing the home locations of the pupils who attend grammar schools we found that the spread of wards in which grammar school pupils live is surprisingly wide. Across England as a whole, one third of the wards in the country (33%) house at least one pupil who attends a grammar school. At the LA level, 80 LAs have more than 1% of the pupils who live in their area attending grammar schools, even though only 36 LAs actually have grammar schools of their own. Nationally, about 20% of grammar school pupils come from outside the LA; for some LAs, this figure is as high as 75%.

From this it is clear that LAs do not form neat, self-contained selective systems. Even talking about a 'selective LA' is problematic, since for some apparently selective LAs more than half their grammar school pupils originate from outside the LA. Certainly any comparison that uses the LA as the unit of analysis ignores this complexity and must be open to serious question.

9.1.3. Subject difficulties

This is a complex and controversial issue and a full exploration goes well beyond the scope of this report. For now we simply state that if large numbers of apparently typical students gain systematically higher grades in one subject than in all their other subjects, then there is at least a *prima facie* case for questioning the equality of the outcome grades. In a context where a tariff of equivalence between outcomes in different types of qualification has been decreed without any clear evidential basis, and there appears to be considerable anecdotal evidence of non-equivalence, then it is appropriate to question whether some routes are harder than others. On this basis, it is clear from the data that there are substantial differences in the difficulties of different subjects at GCSE and across different qualifications at KS4. The points awarded by the Qualifications and Curriculum Agency for different outcomes are not necessarily in good agreement with the patterns of performance seen in the data.

Given this disparity, it seemed important for us to investigate any differences in the types of qualifications taken by different students, and any systematic differences in the entries from different types of schools. This analysis was reported in Section 6.4 (p144).

It is clear that there is a general tendency for pupils in grammar schools to be entered for the more difficult qualifications, though the overall difference is not huge: grammar school students take subjects that are on average about a tenth of a GCSE grade harder (on all KS4 qualifications) than those in other schools. Moreover, this picture is complicated further by the fact that more able pupils in any school are more likely to be entered for harder subjects. Nevertheless, even after adjustment for this tendency, students in grammar schools still enter slightly harder subjects.

Introducing a correction for the difficulty of the subjects taken into the comparison of performance by grammar and non-grammar schools did make a small difference to the result, in favour of the former. For most models the adjustment for subject difficulties made about 0.2 of a grade's difference to the estimate of the grammar school effect (p211).

9.1.4. Creaming

We developed a method for calculating the creaming effect of each grammar school on each non-grammar school, using data on the home locations of individual pupils who attend each type of school. So far as we are aware, this kind of analysis has never been done before.

Overall, the pattern we found is of widespread, low-level creaming. Relatively few schools are heavily creamed: using our analysis only 161 schools lose more than 20% of their potential pupils to grammar schools, and three quarters of these are in just four LAs. About a third (35%) of non-grammar schools lose between 1% and 20%; a further third (32%) lose between 0 and

1%. Across England as a whole, almost three-quarters of all non-grammar schools are affected, if only slightly, by creaming of grammar schools. For example, most schools in London lose a few pupils to grammar schools, even in areas where there are no local grammar schools.

Across the country there is a general tendency for this creaming to come from schools with relatively able pupils. Hence although they may lose some of their most able pupils, the fact that they have more than their 'fair share' of these pupils anyway means that their overall composition is not unbalanced. It is not until the proportion creamed reaches around 10% that there is a clear impact on the number of high ability pupils in a school.

The intention behind looking at the way grammar schools cream pupils from non-grammar schools was to identify areas or groups of schools that were affected by selection. Having established that the LA was not really an appropriate unit of analysis for comparing selective and non-selective systems, we sought some other way of making the distinction.

Our conclusion, however, is that there is no clear cut-off for defining selective areas. So many schools are so slightly creamed that it is hard to say whether they are really affected by selection or not. The effects of selection are so widespread, albeit tailing off to low levels, that we cannot confidently draw lines around its impact. Self-contained selective systems do not exist in England.

9.1.5. Selectivity of selective and non-selective schools

We also developed a method for calculating the 'selectivity' of a school, based on the differences between the pupils who attend it and their neighbours in the same wards who do not. Again, we believe no-one has used national pupil level data in this way before.

Not surprisingly, grammar schools are a good deal more academically selective than non-grammar schools. However, there is some overlap: some non-selective schools are actually more selective than some grammar schools, which does seem surprising. With regard to social selectivity, again not surprisingly, grammar schools are more selective than other schools. In other words, pupils who attend grammar schools are less likely to be eligible for FSM than their neighbours in other schools. Here, though, the overlap is even greater; the most socially selective schools in the country are not grammar schools, but are theoretically 'non-selective'.

This again is a surprising finding. For a school to be 'selective' according to this definition, some kind of discrimination must be taking place, either in choices made by pupils about where to apply, or in choices made by schools about whom to accept. Note that we are not simply saying that there is segregation across schools, with some having more advantaged populations than others. Such segregation is a well known phenomenon and may simply reflect the social segregation of residential neighbourhoods. Our finding suggests that even within the different neighbourhoods that exist, somehow the most socially advantaged children are finding their way to different schools from the more disadvantaged. This sorting cannot be explained by residential segregation.

Further exploration of the mechanisms by which this sorting takes place is really beyond the scope of this report, though it is certainly a question with a good deal of relevance to recent debates about reforms to school admissions

processes provoked by the 2007 Admissions Code (DfES, 2007). However, the phenomenon is also relevant to any discussion of the social bias inherent in grammar school selection.

9.1.6. Social bias in selection

It is clear from the pupil level national data that pupils in grammar schools have a significantly lower rate of eligibility for Free School Meals (2.0%) than those in other schools (13.3%). This difference does not appear to be adequately explained by the higher abilities of grammar school pupils or their tendency to be located in areas with relatively low social disadvantage. Hence it does seem to suggest the existence of some kind of social bias in their selection processes. Of course, 'selection' here would include any differences in the tendencies of different pupils to apply to a grammar school, or to remain in one having been selected, as well as any direct bias in the schools' selection procedures.

Overall, there is some evidence that grammar schools as a whole appear to take less than their 'fair share' of pupils with FSM. If this is true then it could be argued that the existence of grammar schools helps to perpetuate social inequality and hence that anyone who sees the promotion of equality as an important goal for education should want to see grammar schools replaced by comprehensives.

However, if comprehensive schools in the current system are themselves often even more extreme in their social selectivity than grammar schools (as we have suggested in the previous section), then this would not appear to be any kind of solution to the problem. Indeed, if the choice is between 'social selection, with academic selection as a by-product', as appears to be occurring within the comprehensive system, and 'academic selection, with social selection as a by-product' as seems to be the case within grammar schools, then the latter could certainly be seen as more meritocratic.

Moreover, if social selection by grammar schools occurs within a context of widespread social selection by schools in the system as a whole it becomes difficult to untangle cause and effect. The issue is, once again, rather more complex than appears at first sight.

A yet further complication to this already confused issue is that as well as social selection, and hence segregation, operating at the level of the school, there may also be differential social-academic gradients within schools. For example, even if a comprehensive school takes a representative slice of the population, if it then allocates all its socially most advantaged pupils to the top sets and so confers further advantage on them, while consigning those at the bottom of the social pile to the bottom sets, the fact that both groups are in the same school will be no compensation for the fact that social inequality has been compounded. Our finding that pupils with Free School Meals who get in to grammar schools do relatively better (p218) could be re-expressed as saying that FSM is more of a disadvantage if you go to a comprehensive than it is if you go to a grammar school.

9.2. Comparing performance of grammar and non-grammar schools

9.2.1. Evaluating different models for comparing performance

The results presented in Chapter 8 are complex and hard to interpret. It is far from clear what our conclusion should be regarding the effect of grammar school selection on performance. Before returning to this question (in the next section), there are a number of general comments we can make about these results and their sensitivity to different assumptions and models.

Firstly, it is clear that although the simple comparisons of outcomes (Section 8.2.1, p195), and even the comparisons of pupils matched by KS2 results (Section 8.2.3, p203), illustrate the approach, they are really no substitute for the full value-added models (Section 8.3, p209). The latter may be more complex, but they are also better in that they adjust more fully for initial differences between the pupils in different types of schools.

Secondly, the choice of which variables to take account of in a value-added model is clearly an important one. The grammar school effect is seen to decrease as more explanatory variables are added to the models. If we look at the outcome variable average (weighted) point score, the grammar school effect is 0.9 of a grade after adjusting for KS2 levels (in the OLS model – see Table 41). When all pupil level variables have been introduced, we see the grammar school effect reducing to about 0.6 of a GCSE grade. When the school level variables are introduced, the grammar school effect is hardly above zero (0.05 of a grade). In all models, the grammar school effect looks quite substantial on adjustment for a simple measure of prior attainment, but declines considerably, often to close to zero, when further explanatory variables are included.

The particular issue of whether or not (and, if so, how) school level variables, or compositional effects, should be included in any analysis has been discussed in some detail in Section 8.3.2 (p213). Suffice it to say here that this remains an unresolved and difficult question. It is clear, however, that compositional variables can make a substantial difference to our conclusions.

Thirdly, we see that the choice of different types of statistical model also make an important difference. OLS models generally provide slightly lower estimates of the grammar school effect than ML models, and ML models with variable slopes tend to estimate it slightly higher again. The variable slopes model has the further complication that the 'grammar school effect' will be different for pupils with different characteristics; our choice of 'typical' pupil is somewhat arbitrary.

Fourthly, correcting for the imperfect reliability of our explanatory variables can also make a substantial difference to the results. However, it is hard to know what use we can make of this finding. For one thing, we do not know what the 'true' reliability is; 'reliability' can be understood in different ways and we have no good estimates. Given that correlations between KS2 and KS3 scores can be of the order of 0.9, however, we are probably safe in assuming at least this value. Hence, provided we have actual KS2 scores or their equivalent (rather than just broad levels), reliability should not be too much of an issue. For another thing, if the apparent existence of a compositional effect is really just an artefact of unreliability (Harker and

Tymms, 2004), then including these school composition variables may effectively already adjust for any regression to the mean. In this case, making a separate adjustment for unreliability would be to correct twice, and so potentially overcompensate.

Fifthly, the choice of outcome variable can make a significant difference to the conclusion. In comparing the estimates for the grammar school effect on the six GCSE subjects in Table 40, the results vary by as much as a third of a GCSE grade, the grammar school effect being highest for French and mathematics, and lowest for history and double science. Different ways of combining performance across GCSEs can make an appreciable difference to the estimate of the grammar school effect. In particular in Table 42 we see three different ways of adding the total points achieved by a pupil at KS4 (using the total of the DCSF points allocation for all qualifications, capping at the best eight or equivalent, limiting the total to GCSEs only) with more than a whole grade's difference (ie six points) between estimates from the same model for different outcomes in some cases. Models in which GCSE points are corrected to adjust for subject difficulty (Table 41) appear to estimate the grammar school effect around a quarter grade higher than those without this correction.

Sixthly, the choice of comparison group does not seem to matter too much. Grammar schools can most simply be compared with all other schools. If we try to define selective units in order to group grammar schools with their associated 'creamed schools' (whether or not they are officially designated as 'secondary moderns') in order to compare them with non-selective units, we find that the performance of these 'creamed schools' cannot really be distinguished from that of other non-grammar schools. Hence any such grouping serves merely to dilute any grammar school effect, rather than to provide a compensating force of its own. This is perhaps fortunate, given the difficulty we have had in defining appropriate comparison groups earlier in the report. Moreover it is important since it suggests that whatever the grammar school effect, there does not appear to be any compensatory 'secondary modern' effect on performance.

Seventh and finally, the analysis presented in section 8.4 (p220) suggests that despite a the sophisticated statistics and the range of variables for which we have controlled in all these models, we have still failed to remove some unobserved factor that accounts for increased progress in both primary and secondary school phases for those pupils who attended grammar schools during the latter. For each model that appears to show enhanced progress for secondary school pupils in grammar schools, we find on applying that model to the same pupils to estimate their relative progress through primary school that they are similarly advantaged, even before they have set foot in a grammar school.

9.2.2. What is the overall grammar-school effect?

Perhaps frustratingly, the most defensible answer seems to be that we cannot really say. There are too many problems in trying to find an answer: limitations of the data available and sensitivity to arbitrary assumptions mean that it can never be clear precisely what the effect is.

However, we recognise that such academic caution will seem to some to be evasive, and that there is an understandable desire to present a simple answer

to an apparently simple question. With this in mind, we may tentatively climb off the fence and venture a hesitant opinion, albeit caveat laden.

Within the range of models we have explored, there is a range of results. The selective analyst, particularly if he or she wanted to make the evidence fit a predetermined conclusion, could defend estimates of the grammar school effect anywhere between two-and-a-half GCSE grades per subject in favour of grammar schools (Model 2 in Table 44, p214) to a fifth of a grade the other way (for history GCSE with ML variable slopes model, Table 40, p210). In a less comprehensive (or more selective) analysis, in which only one model were presented, any of these results might certainly appear to be the true answer.

If we limit ourselves to what seem to be the most defensible models, we are still left with a range of conclusions. For aggregated performance across all subjects, most of our best estimates seem to be between around three-quarters of a grade per subject in favour of grammar schools at one end, and no real difference at the other. For individual subjects, the grammar school advantage is generally also between zero and three-quarters of a grade, though it looks to be a little less in English literature and history and perhaps a little more in French and maths.

Hence, despite the great variety of results we have found there is a reasonably consistent pattern to the overall conclusion: attending a grammar school may be associated with a small advantage in achievement, probably between zero and three-quarters of a GCSE grade per subject taken. Given the sensitivity of this result to different assumptions and models, together with the statistical margin of error that qualifies it, it does not seem reasonable to try to place it any more accurately than this.

However, it should be pointed out that most statistical orthodoxy in the estimation of school effects favours the inclusion of compositional variables (Teddlie and Reynolds, 2000) and these models have produced estimates at the bottom end of this range. Furthermore, the fact that the models that estimate positive grammar school effects also appear to estimate similar sized 'effects' for the progress made by those same pupils in primary school (section 8.4, p220) may also lend weight to a preference for placing the true grammar school effect at the lower end of this range.

9.2.3. Comparison of our results with those of previous studies

Generally speaking, our results seem to be in good agreement with previous work, so far as there is overlap. Like the majority of existing studies, we find a small benefit from attending a grammar school, after adjustment for other characteristics.

Unlike some previous studies, we have not compared selective systems (such as LAs) with non-selective ones, since we have shown that there is no good unit for such an analysis.

Some of the previous results for subgroups are replicated in our study. For example, we find that FSM pupils who make it to a grammar school appear to gain an extra advantage.

As well as largely confirming previous findings, we would argue that we have substantially strengthened them, by showing that a number of alternative explanations for these phenomena can be discounted. These

include the impact of regression to the mean with unreliably measured variables and the choice of particular models for analysis. We have also made clear how the results of this kind of analysis are sensitive to the inclusion of different kinds of explanatory variables. Problems with the range and quality of data available remain, however.

9.3. Conclusions

9.3.1. Implications for policy

Implications of research results for policy are always hard to divine. Researchers seem often to fall prey to, on the one hand, overstating what they perceive to be implications by going beyond the data, and, on the other, overestimating the relevance of research to the policy making process. Nevertheless, there do appear to be some lessons that can be learnt from this work.

The first is that there does not appear to be a strong evidence-based case for the abolition of grammar schools. Despite considerable efforts and multiple attempts to configure our analyses differently, we have been unable to find any analysis that convincingly shows a negative impact of grammar school attendance on academic achievement.

Second, we have also failed to find any evidence of collateral harm to any other schools, arising from the existence of grammar schools. Overall, schools are just as likely to be performing well, whether or not they are 'creamed' by a grammar school. Hence, on the basis of KS4 performance at least, there do not appear to be strong grounds for abolishing selection as it currently operates. Of course, we must remember that our analyses have been limited to qualifications achieved at KS4 and say nothing about other effects of selection, such as affective outcomes or longer-term life consequences.

Third, although we appear to have largely endorsed the *status quo*, that should not be interpreted as supporting a return to full-scale selection, as existed prior to 1965. The patterns of grammar school attendance and creaming existing today are very different from what might be found in a fully selective system. Furthermore, the arguments for and against selection are wider than the issue of achievement we have addressed.

Fourth, we have identified what seems to be a significant issue of social selectivity occurring across all types of school. While it appears to be relatively straightforward to understand how this can occur for grammar schools, it is rather more puzzling to see it evident in supposedly non-selective schools. It certainly looks as though some schools' admissions processes are acting to exacerbate social segregation, rather than to reduce it. Hence it seems appropriate to look again at the policy context in which this situation has arisen. From a policy point of view it may be less helpful to ask whether one type of school is better than another than to identify features of each that may be positive or negative.

Fifth, we have suggested that grammar schools may not be taking their fair share of poorer pupils. Grammar schools should therefore examine their selection procedures, with a view to making them fairer. Our knowledge of the tests that are widely used in these procedures, and the industry that

surrounds them, indicates that there may be substantial room for improvement here.

Sixth, and finally, we have been limited in our ability to answer key questions about the impact of selection by the availability of high quality data, despite the fact that what is available now is a substantial improvement on what was available until quite recently. Specific remaining limitations include the need for more finely differentiated outcome measures than just grades at GCSE (especially at the top end) and a better socio-economic indicator than Free School Meals status. Policy makers should continue to work to improve the quality of data available for evaluation.

9.3.2. Implications for further research

This report has taken hundreds of hours to produce and now stands at over 100,000 words in length. It may seem somewhat self-serving to call for yet more research, but we believe there are still important gaps in our knowledge which need to be recognized and, if possible, addressed. Two of these are specifically worth identifying.

The first is that we have been able to say nothing about most of the really important outcomes of education. The extent to which individuals have control over their lives, are healthy, happy and socially responsible, for example, may matter far more than the qualifications they achieve at KS4, and it is clear that schooling can play an important role in these outcomes (Bynner and Feinstein, 2005). Even more obviously 'educational' outcomes such as academic achievement post-16 and success in higher education and employment have been excluded from this research, but again probably matter more than the outcomes we have compared. There is a strong case for further research to include these outcomes.

The second is that if we are to answer the question about the impact of selection conclusively, we need better research designs to control bias. All statistical attempts to adjust for the initial differences between those selected by grammar schools and those attending other schools are inevitably imperfect and allow room for doubt about their conclusions. Of course, the methodologically strongest design, the use of random allocation of pupils to different types of school, would probably be judged unethical and is therefore inappropriate. However, given that admission to a grammar school is often based purely on whether a score on a particular test is above or below a given threshold, it would be possible to use a 'regression-discontinuity' design, which is also methodologically strong. Such research would require access to 11-plus scores, matched to later outcomes; if these were available it would enable us to say with much greater confidence what the impact of grammar school attendance was.

9.3.3. Conclusion

We believe we have used the best available national data and applied the best known statistical analyses to them. Nevertheless, limitations of the data and the research design, together with the need for assumptions that are essentially matters of judgement, have led us to a range of results, rather than a single clear finding. Moreover, the range may be even wider than we have found, and there are plausible arguments that our estimates could be biased in either direction.

Having said that, our main finding may be summarized simply: pupils who attend grammar schools appear to achieve between zero and three-quarters of a GCSE grade per subject more than 'similar' pupils in other schools. This finding is broadly consistent with the results of the best previous studies.

Whether or not the pupils who are unsuccessful in applying to grammar schools achieve any less than they would if there were no grammar schools is much harder to say, given the complex patterns of 'creaming' we have reported. However, we have not been able to find any evidence that this is so.

Our evidence will no doubt be welcomed by supporters of grammar schools. We must remember, however, that the debate about selection goes well beyond its impact on academic outcomes, as we have outlined in Chapter 2. It would be quite possible to accept all our findings yet argue logically and consistently against selection.

10. Appendix: Results from all models

In this appendix we present the detailed results of the multilevel models and the OLS (ordinary least squares) regression models discussed in Chapter 8.

The numbers in the tables are the regression coefficients, or slopes of the lines, and can be interpreted as relating to GCSE grades, where the difference between two consecutive grades is 6 points.

The tables show the effect of introducing the various input, or explanatory variables, in different combinations. In the initial analysis 18 models were investigated using OLS. Of these 11 were chosen for a similar analysis using multilevel modelling. A selected number of the OLS models are included, for direct comparison with the multilevel models. Thus in both sets of tables the models are numbered as 1 to 11.

Results are presented for each of the outcome variables considered in Section 8.1.1:

Aggregated points scores:

- Total points score; all GCSEs and their equivalents
- Capped points score; best 8; all GCSEs and their equivalents
- Total points; full and short GCSEs
- Average points score; full and short GCSEs
- Rasch score of overall GCSE achievement

Individual GCSE subject points scores:

- Mathematics
- English
- Double science
- History
- French
- English literature

The number R-squared (or R^2), which appears in the tables, is a measure of the fit of the model. R-squared takes a value between 0 and 1, and in general terms the higher the value, the more the variation in the outcome variable is explained by the model.

10.1. Ordinary Least Squares Models

OLS Regression	KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables		KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables
Outcome	**Total Points**						**Capped Points**				
Equivalent multilevel model	3	5	7	9	11		3	5	7	9	11
Intercept	-181.9	-34.8	-67.1	-6.2	-162.3		-77.6	27.4	7.1	56.2	-48
Pupil level variables											
KS2 average level	135.6		19.7	15.0	14.4		92.2		12.4	8.9	8.3
KS2 average mark		6.8	6.0	5.9	5.8			4.5	4.0	3.9	3.8
Male				-29.2	-29.9					-20.5	-20.2
FSM				-32.7	-35.0					-23.4	-23.9
idaci				-94.6	-108.0					-81.5	-73.3
Ethnicity Category A				7.4	5.7					7.8	6.8
Ethnicity Category B				61.0	56.1					42.2	40.7
School level variables											
School average KS2 level					38.6						27.3
School % FSM					-68.3						-59.7
School % FSM squared					167.1						136
School idaci					68.5						5.6
Single sex school					-0.5						7.0
Boys school					9.9						0.7
R-squared	0.37	0.38	0.38	0.42	0.42		0.4	0.41	0.41	0.46	0.47
Grammar school	47.7	33.1	32.7	25.5	-1.5		38.3	29.5	29.2	23.5	-0.75
Grammar school effect in GCSE grades	**7.9**	**5.5**	**5.4**	**4.3**	**-0.3**		**6.4**	**4.9**	**4.9**	**3.9**	**-0.1**

OLS Regression	KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables		KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables
Outcome	**Total GCSE Points**						**Average GCSE points**				
Equivalent multilevel model	3	5	7	9	11		3	5	7	9	11
Intercept	-194.6	-53.2	-92.5	-14.5	-177.0		-4.7	6.3	3.7	9.2	-2.6
pupil level variables											
KS2 average level	129.7		24.0	18.5	17.5		10.1		1.6	1.2	1.1
KS2 average mark		6.5	5.5	5.3	5.2			0.5	0.4	0.4	0.4
Male				-32.6	-32.1					-2.3	-2.3
FSM				-29.7	-27.5					-2.2	-2.1
idaci				-138.0	-90.7					-9.5	-6.9
Ethnicity Category A				12.5	11.7					1.0	0.9
Ethnicity Category B				52.5	55.8					4.1	4.1
school level variables											
School average KS2 level					45.2						3.2
School % FSM					-121.9						-7.1
School % FSM squared					223.7						16
School idaci					-58.8						-3.5
Single sex school					15.9						1.4
Boys school					3.0						-0.4
R-squared	0.43	0.45	0.45	0.51	0.52		0.46	0.48	0.48	0.54	0.55
Grammar school	73.8	60.3	59.7	51.3	3		5.2	4.2	4.1	3.5	0.3
Grammar school effect in GCSE grades	**12.3**	**10.05**	**10**	**8.55**	**0.5**		**0.9**	**0.7**	**0.7**	**0.6**	**0.1**

OLS Regression	KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables
Outcome	**Rasch GCSE Score**				
Equivalent multilevel model					
Intercept	-0.2	9.6	6.8	11.1	-0.4
pupil level variables					
KS2 average level	9.2		1.7	1.4	1.3
KS2 average mark		0.5	0.4	0.4	0.4
Male				-2.1	-2.0
FSM				-1.5	-1.4
idaci				-7.7	-5.5
Ethnicity Category A				1.0	0.9
Ethnicity Category B				3.5	3.5
school level variables					
School average KS2 level					3.1
School % FSM					-5.7
School % FSM squared					13.2
School idaci					-3.0
Single sex school					1.3
Boys school					-0.5
R-squared	0.51	0.54	0.54	0.59	0.6
Grammar school	6.2	5.2	5.2	4.7	1.65
Grammar school effect in GCSE grades	**1**	**0.9**	**0.9**	**0.6**	**0.3**

OLS Regression	KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables		KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables
Outcome	**GCSE Mathematics**						**GCSE English**				
Equivalent multilevel model	3	5	7	9	11		3	5	7	9	11
Intercept	-9.2	2.8	3.2	6.3	-0.6		-0.3	11.1	7.1	12.3	1.58
pupil level variables											
KS2 average level	11.3		-0.3	-0.4	-0.4		9.5		2.5	2.0	2.0
KS2 average mark		0.6	0.6	0.6	0.5			0.5	0.4	0.4	0.3
Male				-0.1	0.0					-3.3	-3.2
FSM				-1.7	-1.7					-1.9	-1.8
idaci				-8.3	-5.7					-7.7	-5.6
Ethnicity Category A				0.9	0.9					1.0	1
Ethnicity Category B				4.3	4.4					3.3	3.3
school level variables											
School average KS2 level					2.0						2.9
School % FSM					-7.5						-5.2
School % FSM squared					15.4						11.9
School idaci					-4.0						-2.7
Single sex school					1.5						1.1
Boys school					-0.6						0.2
R-squared	0.45	0.49	0.49	0.51	0.52		0.42	3.8	0.43	0.48	0.49
Grammar school	5.2	3.9	3.9	3.4	0.9		4.5	3.8	3.7	3.2	0.3
Grammar school effect in GCSE grades	**0.9**	**0.65**	**0.66**	**0.6**	**0.2**		**0.8**	**0.6**	**0.6**	**0.53**	**0**

OLS Regression	KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables		KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables
Outcome	**GCSE Double Science**						**GCSE History**				
Equivalent multilevel model	3	5	7	9	11		3	5	7	9	11
Intercept	-0.3	4.6	3.0	6.0	-2.2		-14.0	0.7	-2.4	12.3	-10.4
pupil level variables											
KS2 average level	10.8		1.0	1.0	0.9		12.3		1.9	2.0	1.6
KS2 average mark		0.5	0.5	0.5	0.5			0.6	0.5	0.4	0.5
Male				-0.1	0.2					-3.3	-1.7
FSM				-2.0	-1.9					-1.9	-2.2
idaci				-8.9	-6.0					-7.7	-7.2
Ethnicity Category A				0.8	0.7					1.0	0.9
Ethnicity Category B				3.6	3.8					3.3	4
school level variables											
School average KS2 level					2.3						3.5
School % FSM					-6.7						-3.2
School % FSM squared					14.5						11.8
School idaci					-5.1						-7.2
Single sex school					1.8						1.8
Boys school					-1.9						-0.9
R-squared	0.4	0.43	0.43	0.46			0.36	0.39	0.39	0.48	0.42
Grammar school	4.5	3.6	3.6	3.1			3.8	2.9	2.9	3.2	-1.1
Grammar school effect in GCSE grades	**0.7**	**0.6**	**0.6**	**0.5**			**0.6**	**0.5**	**0.5**	**0.5**	**-0.2**

OLS Regression	KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables		KS2 average level	KS2 average mark	KS2 average mark and level	Pupil level variables	Pupil level variables and school level variables
Outcome	**GCSE French**						**GCSE English Literature**				
Equivalent multilevel model	3	5	7	9	11		3	5	7	9	11
Intercept	-6.0	6.2	3.0	6.7	-3.6		-1.2	10.8	7.2	12.2	-0.6
pupil level variables											
KS2 average level	10.3		2.0	1.6	1.5		9.8		2.3	1.8	1.7
KS2 average mark		0.5	0.4	0.4	0.4			0.5	0.4	0.4	0.4
Male				-3.7	-3.9					-3.6	-3.5
FSM				-1.3	-1.3					-1.9	-1.8
idaci				-6.9	-5.1					-8.0	-5.8
Ethnicity Category A				1.6	1.4					1.2	1.1
Ethnicity Category B				2.8	2.9					3.6	3.7
school level variables											
School average KS2 level					2.8						3.4
School % FSM					-9.1						-2.7
School % FSM squared					20.1						7.9
School idaci					-1.5						-3.3
Single sex school					0.6						1.2
Boys school					1.2						0.1
R-squared	0.37	0.39	0.39	0.44	0.45		0.3	0.34	0.34	0.4	0.41
Grammar school	4.8	4.2	4.1	4	1.1		3.7	3.1	3	2.5	-0.7
Grammar school effect in GCSE grades	**0.8**	**0.7**	**0.7**	**0.7**	**0.2**		**0.6**	**0.5**	**0.5**	**0.4**	**-0.1**

10.2. Multilevel Models

10.3. KS1 to KS4 models for Section 8.4

Table 47: OLS regression results for KS2 to KS4 showing the selective systems effects in effect sizes

Table 48: OLS regression results for KS1 to KS2 showing the selective systems effects in effect sizes

Table 49: OLS regression results for KS2 to KS4 showing the grammar schools effects in effect sizes

Table 50: OLS regression results for KS1 to KS2 showing the grammar school effects in effect sizes

10.4. Results from Propensity Score Matching, Section 8.5

Table 51: Output from logistic regression model for calculation of propensities

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	62854.411	.192	.665

Classification Table(a)

Observed			Predicted		
			Grammar Dummy		Percentage Correct
			.00	1.00	.00
Step 1	Grammar Dummy	.00	470907	4182	99.1
		1.00	8266	11829	58.9
	Overall Percentage				97.5

a The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	FSMDummy	-.779	.063	150.812	1	.000	.459
	whiteDummy	-.451	.031	206.141	1	.000	.637
	indianDummy	.924	.058	253.999	1	.000	2.519
	idaci_06	1.290	.087	219.810	1	.000	3.634
	KS2AveLevel	-.225	.066	11.710	1	.001	.799
	KS2MarkEng	.050	.002	654.398	1	.000	1.051
	KS2MarkMath	.079	.002	1937.606	1	.000	1.082
	KS2MarkSci	-.012	.002	24.682	1	.000	.988
	KS2TeachAEng	.726	.028	666.020	1	.000	2.067
	KS2TeachAMath	.696	.031	508.590	1	.000	2.005
	KS2TeachASci	.223	.030	54.551	1	.000	1.249
	wardpcgrammar	.911	.793	1.320	1	.251	2.486
	wardxks2	.607	.011	2859.335	1	.000	1.836
	int2	-2.153	.040	2917.516	1	.000	.116
	int3	.255	.007	1403.179	1	.000	1.290
	Constant	-20.995	.254	6825.721	1	.000	.000

a Variable(s) entered on step 1: FSMDummy, whiteDummy, indianDummy, idaci_06, KS2AveLevel, KS2MarkEng, KS2MarkMath, KS2MarkSci, KS2TeachAEng, KS2TeachAMath, KS2TeachASci, wardpcgrammar, wardxks2, int2, int3.

Table 52: Group means for propensity-matched sub-sample

	School type	N	Mean	Std. Deviation	Std. Error Mean
Student Sex Dummy	Non-grammar	14115	.4801	.49962	.00421
	Grammar	14115	.4903	.49992	.00421
White Dummy	Non-grammar	14115	.8045	.39657	.00334
	Grammar	14115	.8168	.38685	.00326
Pakistani Dummy	Non-grammar	14115	.0116	.10700	.00090
	Grammar	14115	.0189	.13623	.00115
Indian Dummy	Non-grammar	14115	.0447	.20674	.00174
	Grammar	14115	.0436	.20430	.00172
Other White Dummy	Non-grammar	14115	.0396	.19512	.00164
	Grammar	14115	.0236	.15178	.00128
African Dummy	Non-grammar	14115	.0057	.07554	.00064
	Grammar	14115	.0058	.07600	.00064
Caribbean Dummy	Non-grammar	14115	.0078	.08814	.00074
	Grammar	14115	.0062	.07872	.00066
Bangladeshi Dummy	Non-grammar	14115	.0027	.05216	.00044
	Grammar	14115	.0033	.05700	.00048
Chinese Dummy	Non-grammar	14115	.0043	.06506	.00055
	Grammar	14115	.0122	.10972	.00092
Free School Meals status	Non-grammar	14115	.0269	.16186	.00136
	Grammar	14115	.0270	.16207	.00136
IDACI deprivation index	Non-grammar	14115	.1324802	.12733135	.00107175
	Grammar	14115	.1320070	.12324762	.00103738
Average KS2 mark	Non-grammar	14115	75.0147	6.59240	.05549
	Grammar	14115	75.0712	6.59185	.05548
Average KS2 Teacher Assessment	Non-grammar	14115	4.7384	.37559	.00316
	Grammar	14115	4.7499	.35801	.00301
Average GCSE score (full, short and vocational GCSEs only)	Non-grammar	14067	45.0092	7.04477	.05940
	Grammar	14101	48.3259	5.44815	.04588
Rasch scores of overall GCSE achievement	Non-grammar	14054	45.2882	7.33207	.06185
	Grammar	14099	49.0872	6.67355	.05620
Predicted probability	Non-grammar	14115	.4218606	.28025625	.00235893
	Grammar	14115	.4218606	.28025617	.00235893

Table 53: T-tests for differences on key variables between grammar and non-grammar school pupils in propensity-matched samples

		F	[illegible]	t	[illegible]				[illegible]	[illegible]
[illegible]	[illegible]		[illegible]		[illegible]	[illegible]	[illegible]		0	[illegible]
	[illegible]				7		[illegible]			
[illegible]	[illegible]			2	8	[illegible]				
	[illegible]				[illegible]	[illegible]				
[illegible]	[illegible]		[illegible]			[illegible]				
	[illegible]				[illegible]	[illegible]				
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]				
	[illegible]			[illegible]	[illegible]	[illegible]				
[illegible]	[illegible]		[illegible]			[illegible]				
	[illegible]				[illegible]	[illegible]	0		[illegible]	[illegible]
[illegible]	[illegible]									
	[illegible]				[illegible]					
[illegible]	[illegible]									
	[illegible]				[illegible]					
[illegible]	[illegible]									
	[illegible]				[illegible]					7
[illegible]	[illegible]									
	[illegible]				[illegible]					
[illegible]	[illegible]									
	[illegible]				[illegible]					
[illegible]	[illegible]						[illegible]	[illegible]	0	[illegible]
	[illegible]				[illegible]		[illegible]	[illegible]	0	[illegible]
[illegible]	[illegible]									
	[illegible]				0					
[illegible]	[illegible]									
	[illegible]				[illegible]					
[illegible]	[illegible]						5		2	9
	[illegible]				[illegible]		5		6	4
[illegible]	[illegible]						7		5	9
	[illegible]				[illegible]		7		7	6
[illegible]	[illegible]						[illegible]	[illegible]	9	[illegible]
	[illegible]				0		[illegible]	[illegible]	9	[illegible]

11. References

Ammermueller, A. (2005) Educational Opportunities and the Role of Institutions, *ZEW Discussion Paper*, No. 05-44.

Atkinson, A., Gregg, P and McConnell, B. (2004) *The Result of 11 Plus Selection: An Investigation into Equity and Efficiency of Outcomes for Pupils in Selective LEAs.* Available online at: http://www.bris.ac.uk/Depts/CMPO/workingpapers/greggwp/greggrammar.pdf

Atkinson, A., Gregg, P. and McConnell, B. (2006) *The Result of 11 Plus Selection: An Investigation into Opportunities and Outcomes for Pupils in Selective LEAs.* Available online at: http://www.bristol.ac.uk/cmpo/workingpapers/wp150.pdf

Blanden, J., Gregg, P. and Machin, S. (2005) *Intergenerational Mobility in Europe and North America.* London School of Economics / Sutton Trust.

Bynner, J. and Feinstein, L. (2005) 'What can policy learn from the research on the wider benefits of learning?' *London Review of Education*, 3, 3, 177-190.

BBC (2005) 'Failing' schools rise to the top. BBC News, 7 December 2005, http://news.bbc.co.uk/1/hi/education/4503276.stm

BBC (2006) 'Richer pupils' at church schools. BBC News, 12 February 2006 http://news.bbc.co.uk/1/hi/education/4707452.stm

BBC (2007a) School admissions 'must be fair'. BBC News, 1 November 2007 http://news.bbc.co.uk/1/hi/education/7073906.stm

BBC (2007b) Shadow education secretary David Willett's speech to the CBI on the Conservatives' backing for Labour's academy system. BBC News, 16 May 2007 http://news.bbc.co.uk/1/hi/uk_politics/6662219.stm

Blanden, J., Gregg, P. & Machin, S. (2005) *Intergenerational Mobility in Europe and North America*. Centre for Economic Performance.

Burgess, S., McConnel, B., Propper, C. & Wilson, D. (2004) *Sorting and Choice in English Secondary Schools*. CMPO Working Paper Series No. 04/111.

Burgess, S., Propper, C. & Wilson, D. (2005) Will More Choice Improve Outcomes in Education and Health Care? The Evidence from Economic Research. The Centre for Market and Public Organisation.

Campbell, D.T. and Boruch, R.F. (1975) Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects, in C.A.Bennett and A.A. Lumsdaine (Eds) *Evaluation and Experiment: Some critical issues in assessing social programs*. New York: Academic Press

Campbell, D.T., and Stanley, J.C. (1963) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Clark, D. (2007a). Selective Schools and Academic Achievement (under review). Available online at: http://bear.cba.ufl.edu/clarkd/research/selective.pdf.

Clark, D. (2007b). Selection versus Comprehensives: Which delivers the best educational outcomes? *Research in Public Policy*, 5, p14-17.

Coe, R. (2002) *It's the effect size, stupid: what effect size is and why it is important.* Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 12-14 September 2002. [Available from http://www.leeds.ac.uk/educol/]

Coe, R., Searle, J., Barmby, P., Jones, K. and Higgins, S. (2008) *Relative difficulty of examinations in different subjects*. Report for SCORE (Science Community Supporting Education), July, 2008. Available at http://www.cemcentre.org/documents/news/SCORE2008report.pdf

Critchlow, J. and Coe, R. (2003) *Serious flaws arising from the use of the median in calculating value-added measures for UK school performance tables.* International Association for Educational Asssessment (IAEA) Annual Conference, October 2003. www.aqa.org.uk/support/iaea/papers/critchlow-coe.pdf

Crook, D., Powers, S. and Whitty, G. (1999) The Grammar School Question. A Review of research on comprehensive and selective education. *Institute of Education, University of London*

Croxford,L. (2000) Inequality in Attainment at Age 16: A 'Home International' Comparison, *CES Briefing No 19*, Edinburgh: Centre for Educational Sociology, University of Edinburgh (May).

Croxford, L. & Paterson, L. (2006) Trends in Social Class Segregation between Schools in England, Wales and Scotland since 1984. *Research Papers in Education*, 21(4), 381-406.

Cuttance, P. (1994) Monitoring educational quality through performance indicators for schools practice. *School Effectiveness and School Improvement*, 5, p101-126.

Daly, P. and Shuttleworth, I. (2000). *After Post-compulsory Education: First Destinations of Northern Ireland Year 12 pupils in 1999.* (Bangor: Department of Education for Northern Ireland).

DfES (Department for Education and Skills) (2007) *School Admissions Code of Practice*. London: DfES. Available at http://www.dfes.gov.uk/sacode/docs/DfES%20Schools%20text%20final.pdf

DfES (Department for Education and Skills) (2005) *Higher standards, better schools for all: More choice for parents and pupils.* Norwich: HMSO. Available at http://www.dfes.gov.uk/publications/schoolswhitepaper

DfES (Department for Education and Skills) (2006) *Contextual Value Added (CVA) Information*. Available at http://www.standards.dfes.gov.uk/performance/1316367/CVAinPAT2005/?version=1 [Accessed 19.6.06]

Education Guardian (2005) Looking out for number one. Education Guardian, 22 February 2005 http://education.guardian.co.uk/schools/story/0,5500,1419398,00.html

Fitz-Gibbon, C. T. and Vincent, L. (1994) *Candidates' Performance in Public Examination in Mathematics and Science*, London: School Curriculum and Assessment Authority.

Fitz-Gibbon, C. T. and Vincent, L. (1997) Difficulties regarding subject difficulties: developing reasonable explanations for observable data. *Oxford Review of Education,* 23 (3) p291–8.

Fogelman, K. (1984) Problems in Comparing Examination Attainment in Selective and Comprehensive Secondary Schools. *Oxford Review of Education,* 10 (1) p33-43.

Galindo-Rueda and Vignoles (2004) The Heterogeneous Effect of Selection in Secondary schools: Understanding the Changing Role of Ability. *IZA Discussion Paper Series*, no 1245, August, 2004. Available at: http://opus.zbw-kiel.de/voltexte/2005/2434/pdf/dp1245.pdf

Gallagher, T. and McKeown, E. (2000) *Attitudes to Education.* (Bangor: Department of Education for Northern Ireland).

Gallagher, T. and Smith, A. (2000) *The Effects of the Selection System of Secondary Education in Northern Ireland: Main Report.* (Bangor: Department of Education for Northern Ireland).

Galton, M. (2002) Continuity and progression in science Teaching at Key Stage 2 and 3. *Cambridge Journal of Education*, 32 (2) p249-265

Gardner, H. (1983) *Frames of Mind: The theory of multiple intelligences.* New York: Basic Books.

Gardner, J. and Cowan, P. (2005) "The fallibility of high stakes '11-plus' testing in Northern Ireland", *Assessment in Education: Principles Policy and Practice*, 12, 2, p145-165.

Goldstein, H. (1984) The Methodology of School Comparisons. *Oxford Review of Education*, 10 (1) p69-74.

Goldstein, H. (1995) *Multilevel Statistical Models*. London: Edward Arnold, New York: Wiley.

Goldstein, H. and Cresswell, M. (1996) The comparability of different subjects in public examinations: a theoretical and practical critique, *Oxford Review of Education*, vol. 22, no. 4, pp. 435–42.

Gorard S. (1999) "Well. That about wraps it up for school choice research": a state of the art review. *School Leadership and Management*, Vol.19,no.1: Feb 99 p25-47

Gorard, S. (2003) What is Multi–level Modelling For?. *British Journal of Educational Studies* 51 (1), 46-63.

Gorard, S., Taylor, C. and Fitz, J. (2003) *Schools, markets and choice policies*. London: RoutledgeFalmer.

Gray, J. and Jesson, D. (1989) The impact of comprehensive reforms. In Lowe R. (Ed) *The changing secondary school*. London, Falmer Press

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., Goldstein, H. & Jesson, D. (1996) Changes and improvements in schools' effectiveness: trends over five years. *Research Papers in Education*, 11 (1) p35-51

Griggs, C. (1989) The new right and English secondary education. In Lowe R. (Ed) *The changing secondary school*. London, Falmer Press

Hanushek, E.A. & Wößmann, L. (2006) Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *The Economic Journal*, 116, C63-C76.

Harker, R. and Tymms, P. (2004) 'The effects of student composition on school outcomes'. *School Effectiveness and School Improvement*, 15, 2, p177-199.

House of Lords (2000) This Department for Education and Employment study was reported in a Parliamentary Answer, House of Lords 6 April 2000, Column WA152. Available at: http://www.publications.parliament.uk/cgi-bin/ukparl_hl?DB=ukparl&STEMMER=en&WORDS=J0bach+perform+&COLOUR=Red&STYLE=s&URL=/pa/ld199900/ldhansrd/vo000406/text/00406w02.htm#00406w02_spnew2

Institute for Public Policy Research (2007) School admissions should offer parents fairer choices. Ippr, 26 February 2007 http://www.ippr.org/pressreleases/archive.asp?id=2588

Jenkins, S.P., Micklewright, J. & Schnepf, S.V. (2006) Social Segregation in Secondary Schools: How Does England Compare with Other Countries? *ISER Working Paper 2006-2*. Colchester: University of Essex. Available online at: http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2006-02.pdf

Jesson, D. (2000) The Comparative Evaluation of GCSE Value-Added Performance by Type of School and LEA. *Discussion Papers in Economics, No. 2000/52*, Centre for Performance Evaluation and Resource Management.

Jesson, D. (2001) Selective systems of education - blueprint for lower standards? *Education Review,* 15 (1), p8-14.

Jesson, D. (2007) A ladder of opportunity? The pupil intake and performance of England's grammar schools. *Research in Public Policy*, 5, p11-13.

Kelly, A. (1976) A study of the comparability of external examinations in different subjects. *Research in education*, 16, p50-53.

Kelly, A. (1976) A study of the comparability of external examinations in different subjects. *Scottish educational studies*, vol 8 no 1, p36-63.

Kerchoff,A.C., Fogelman,K., Crook,D. and Reeder, D. (1996) *Going comprehensive in England and Wales: a study of uneven change*. London: Woburn Press

Levačić & Marsh, A.J. (2007) Secondary modern schools: are their pupils disadvantaged? *British Educational Research Journal*, *33*(2), 155-178.

Manning, A. and Pischke, J. (2004) Ability tracking and student performance in secondary schools in England and Wales, NBER's program on the *Economics of Education* March 2004, National Bureau of Economic Research and London School of Economics.

Manning, A. & Pischke, J. (2006) *Comprehensive versus Selective Schooling in England and Wales: What Do We Know?* IZA DP No. 2072. Institute for the

Study of Labor. Available online at: http://cee.lse.ac.uk/cee%20dps/ceedp66.pdf

Marks, J. and Cox, C. (1984) Educational Attainment in Secondary Schools, *Oxford Review of Education, 10: 1*, p7-31.

Marks, J., and Pomian-Srzednicki, M. (1985) *Standards in English Schools: Report no 2.* London: National Council for Educational Standards.

Marks, J., Cox, C., and Pomian-Srzednicki, M. (1983) *Standards in English Schools: Report no 1.* London: National Council for Educational Standards.

Marsh, H.W. (1991) The failure of high-ability high schools to deliver academic benefits: the importance of academic self-concept and educational aspirations. *American Educational Research Journal,* 28, p445-480.

Maurin, E. & McNally, S. (2007) Educational Effects of Widening Access to the Academic Track: A Natural Experiment. Centre for the Economics of Education: London School of Economics. Discussion Paper No.85. Available online at: http://cee.lse.ac.uk/cee%20dps/ceedp85.pdf

McCulloch, G. (2002) Secondary Education, In Aldridge R (Ed) *A Century of Education*. Routledge & Falmer

Moffitt, T.E., Caspi, A., Harkness, A.R. and Silva, P.A. (1993) 'The natural history of change in intellectal performance: Who changes? How much? Is it meaningful?' *Journal of Child Psychology and Psychiatry*, 34, p455-156.

Mosteller, F., Light, R.J. and Sachs, J.A. (1996) 'Sustained inquiry in education: lessons from skill grouping and class size.' *Harvard Educational Review*, 66, p797-842.

Nash, R. (2003) 'Is the School Composition Effect Real? A Discussion With Evidence From the UK PISA Data'. *School Effectiveness and School Improvement*, 14, 4, p441-457.

Norwood Report (1943) 'Curriculum and Examinations in Secondary Schools: Report of the Committee of the Secondary School Examination Council Appointed by the Board of Education 1941.' Published 1943 HMSO.

Prais, S.J. (2001). Grammar Schools' Achievements and the DfEE's Measures of Value-added: and attempt at clarification. *Oxford Review of Education,* 27 (1) p69-73.

Rosenthal, R. (1966) *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.

Schagen, I. and Schagen, S. (2003) Analysis of National Value-Added Datasets to Assess the Impact of Selection on Pupil Performance. *British Educational Research Journal,* 29 (4) p561-582.

Schagen, I. and Schagen, S. (2005) Combined multi-level analysis with national value-added data sets – a case study to explore the effects of diversity. *British Educational Research Journal.* 31 (3) p309-328

Shuttleworth, I. and Daly, P. (2000) *The Pattern of Performance at GCSE.* (Bangor: Department of Education for Northern Ireland).

Sparkes, B. (2000) Subject comparisons: A Scottish perspective. *Oxford Review of Education*, 26, 2, 175-189.

Steedman, J. (1980) *Progress in Secondary Schools*. London: National Children's Bureau.

Steedman, J. (1983) *Examination Results in Selective and Nonselective Schools: Findings from the National Child Development Study*. London: National Children's Bureau.

Sternberg, R. J. (1996) *Successful intelligence*. New York: Simon & Schuster.

Sullivan, A. and Heath, A. (2002) State and Private Schools in England and Wales. *Sociology working papers* (paper number 2002-02). Available at: http://www/sociology.ox.ac.uk/swps/2002-02.html

Teddlie, C., and Reynolds, D. (2000) (Ed.s) *The International Handbook of School Effectiveness Research*. London: Routledge/Falmer.

Teddlie, C., Reynolds, D. and Sammons, P. (2000) 'The methodology and scientific propreties of school effectiveness research' in C. Teddlie and D. Reynolds (Ed.s) *The International Handbook of School Effectiveness Research*. London: Routledge/Falmer.

Teddlie, C., Stringfield, S. and Reynolds, D. (2000) 'Context issues within school effectiveness research' in C. Teddlie and D. Reynolds (Ed.s) *The International Handbook of School Effectiveness Research*. London: Routledge/Falmer.

Waldinger, F. (2007) *Does Ability Tracking Exacerbate the Role of Family Background for Students' Test Scores?*, Unpublished Manuscript, London School of Economics. Available online at: http://personal.lse.ac.uk/waldinge/Does%20Tracking%20Exacerbate%20the%20Importance%20of%20Family%20Background.pdf

West, A., Hind, A. and Pennell, H (2004) School admission and selection in comprehensive school: policy and practice. *Oxford Review of Education*, 30 (3) p347 -369

West, A. & Hind, A. (2006). Selectivity, admissions and intakes to 'comprehensive' schools in London, England. *Educational Studies*, 32(2), 145-155.

Whitty, G., Powers, S. and Whitty, G. Devolution of choice in education; the school, the state and the market. OU Press

Wiliam, D. (2001) What is wrong with our educational assessments and what can be done about it. *Education Review*, 15 (1) p57-62

Yang, M., and Woodhouse, G. (2001) Progress from GCSE to A and AS level: institutional and gender differences, and trends over time. *British Journal of Educational Research*, 27 (3): 245-267.