Subject Choice and Attainment: State Secondary Schools in Scotland

Anne S Gasteen
University of Stirling

A thesis submitted to the Division of Economics of the University of Stirling in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2018

Declaration

In accordance with the Regulations for Higher Degrees by Research, I hereby declare that the whole thesis now submitted for the candidature of Doctor of Philosophy is a result of my own research and independent work except where reference is made to published literature. I also hereby certify that the work embodied in this thesis has not already been submitted in any substance for any other degree and is not being concurrently submitted in candidature for any degree from any other institute of higher learning. I am responsible for any errors and omissions present in the thesis.

Candidate:

____

Anne S Gasteen

Acknowledgements

Firstly, thank you to my partner, John, for his encouragement, support and understanding throughout but particularly in those last few, very fraught, days when I thought I wouldn't get there. It really would not have been possible without you.

Thank you to my supervisor, Professor David Bell, for reading early drafts of papers, thoughtful comments, encouragement and, not least, his patience. Thanks are due also to my various Heads of Department for their support over the years, with particular mention of Ailsa McKay at the beginning and Stuart McChlery towards the end; your support was much appreciated.

Thank you to my son and daughter, Andrew and Catherine, for their support. I am immensely proud of the people you are.

Thank you also to various friends who encouraged me to 'keep going', 'write some stuff' and for the tales from the PhD bunker, with particular mention of Carolyn Davidson, who first drew my attention to the largely unexploited data a number of years ago.

Audere est facere

"We were working class and no one cared."
(Comment of a friend on our 1970s, North London, Comprehensive school education.)

For my mother and grandmother, greatly missed.

Abstract

Secondary education in Scotland is characterised by substantial socio-economic inequalities in attainment and gendered patterns of performance. Individuals from the most deprived backgrounds do significantly and systematically less well than those from more affluent households while boys underachieve compared to girls. Evaluating attainment in terms of numbers of qualifications achieved, ignores the importance of subject choice. Some subjects are more important than others for progression to tertiary education and employment opportunities. This thesis exploits Scottish Qualifications Authority administrative data, from 2002 to 2009 for state secondary schools, to investigate subject choice and attainment in *facilitating* subjects; traditional academic subjects that facilitate university entry. Chapter One uses sequential logit analysis to examine the decision to stay on at school to take Highers (qualifications necessary for university access) and the decision to take four or more Highers in facilitating subjects (the crucial number for entry to prestigious universities). Chapter Two employs multinomial logit analysis to examine attainment in individual facilitating subjects. Chapter Three uses logistic regression in the context of the Twin Testosterone Transfer hypothesis to explore whether gendered patterns of choice and attainment in Maths and Science might have a biological component in terms of increased testosterone exposure. Despite being in the top 50% for academic achievement nationally, individuals from the most deprived 20% of households were found to be 26% less likely to study four or more facilitating Highers compared with the most affluent 20%. Once facilitating subjects have been chosen, children's ability was seen to be important for securing a low pass at Higher but insufficient to overcome socio-economic disadvantage to achieve the higher grades required by more prestigious universities. There was no evidence of any biological testosterone effect to explain gendered subject choice and attainment patterns. Stark socio-economic background effects revealed a fundamental social inclusion problem with respect to STEM education in Scottish secondary schools.

Table of Contents

List of Tables

List of Figures

Appendix A

List of Abbreviations

AH	Advanced Higher qualifications
AME	Average marginal effect
CfE	Curriculum for Excellence
CSE	Certificate of Secondary Education
CSS	Closely spaced siblings
FE	Further Education
FSM	Free School Meals
GCSE	General Certificate of Secondary Education
H	Higher
HE	Higher Education
LA	Local Authority
LFS	Labour Force Survey
NCDS	National Child Development Survey
NQ	National Qualification
OECD	Organisation for Economic Cooperation & Development
OVB	Omitted variable bias
PISA	Programme for International Student Assessment
S4 / S5 / S6	Senior school year: 4 (age 15-16), 5 (age 16-17), 6 (age 17-18)
SCQF	Scottish Credit & Qualifications Framework
SG	Standard Grade
SGI	Standard Grade & Intermediate 2 qualifications
SIMD	Scottish Index of Multiple Deprivation
SQA	Scottish Qualifications Authority
STEM	Science, Technology, Engineering & Maths
TTT	Twin Testosterone Transfer Hypothesis
UCAS	Universities and Colleges Admissions Service
YCS	Youth Cohort Studies

Introduction

Education is of interest to Economists at both the micro- and macroeconomic level for both efficiency and equity reasons. At the microeconomic level, educational attainment is the main determinant of individuals' employment and earnings' opportunities. At the macroeconomic level, the population's educational attainment determines the quality of the labour resource available in an economy and, therefore, may either stimulate or hinder economic growth. This macroeconomic premium, derived from the individuals' investment in their education, means that the social benefits of education are greater than the (individual) private benefits and provides the underlying rationale for state funding of education. The level and type of education are increasingly important as economic activity becomes ever-more globalised and technologically driven. If access to educational opportunities is inequitable, whether because of class, gender or race, then the production of an educated workforce is likely to be inefficient. The innate ability and talents of individuals will not be realised fully within the education system and resultant economic growth will be lower. Educational equality of opportunity enables social mobility and, in so doing, enhances societal cohesion, both of which are key to economic growth.

This thesis examines subject choice and attainment in state secondary schools in Scotland from 2002 to 2009. Scottish Qualifications Authority (SQA) administrative data are used to investigate the impact of gender and socio-economic background effects on subject choice and attainment therein. Secondary school attainment is pivotal in its importance; the subjects studied and grades obtained at this level determine access to tertiary education institutions and courses that, in turn, determine entry to more highly paid occupations. Effectively, the educational choices made by individuals at secondary school are their initial human capital investment decisions as conceptualised by Becker (1964). These decisions, however, are taken in the context of imperfect information, income inequality and differentiated cultural and social capital.

To analyse educational attainment properly and inform policy, it is important to distinguish between the *distribution* and *allocation* of education as argued by Mare (1981); that is between provision and uptake. Reviewing the quality and equity of outcomes in the Scottish education system, the OECD found that there was little variation in quality of provision between schools but attainment varied considerably with better levels of attainment seen in schools in more affluent areas compared to those in more disadvantaged areas (OECD, 2007).

> "Who you are in Scotland is far more important than what school you attend, so far as achievement differences on international tests are concerned. Socio-economic status is the most important difference between individuals." (*ibid*, p15)

More recently, the Commission on School Reform (2013) reported that whilst the quality of provision in Scottish schools was relatively uniform, delivering consistent, good performance, the educational achievement of the most disadvantaged had not been raised. Provision would appear to be equitable but impact is not. The key objective of current Scottish Government education policy is to close the attainment gap between children from the most disadvantaged households and those from the least deprived (or most affluent) households (Scottish Government, 2017a).

The Scottish Education System

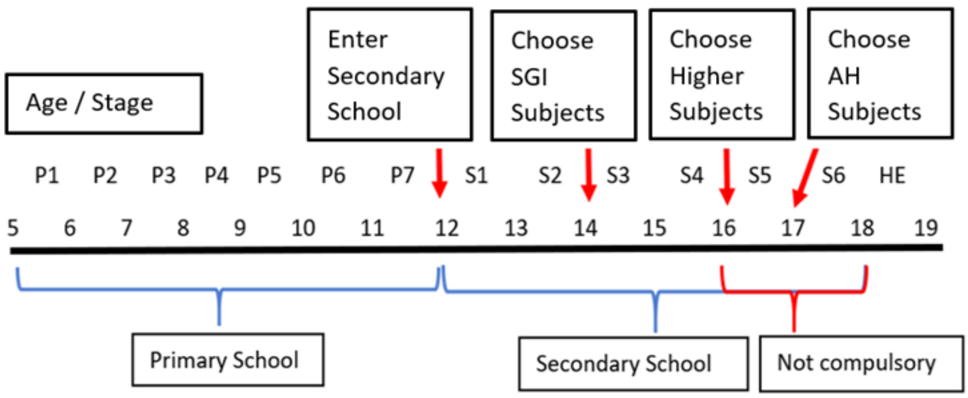
The direction of the state secondary school curriculum is determined by government policy in the first instance with local authorities (LAs) being responsible for its delivery through schools. The timeline of the Scottish educational system for the study period is shown in Figure I.1. This depicts an *on-time*, sequential pattern of progression from one level of study to the next, whereby, pupils studied for the particular qualifications assigned to the different secondary school years. In practice, the system is less linear than this as discussed further below. Table I.1 lists the relevant Scottish secondary school qualifications for the study period. As shown in Figure I.1, Scottish school pupils start secondary school at age 12 (year S1) and make choices at age 14 (year S3) as to which subjects they will study for the following two years for their age-16, formal qualifications at the end of compulsory schooling (year S4). [1] The vast majority of pupils in Scotland now stay on at school beyond the end of compulsory education. Only 11% of S4 pupils left

[1] Scottish school pupils start secondary school a year later than those in the rest of the UK (who start at age 11) but choose their subjects for formal age-16 qualifications at the same time.

school at this stage in 2015/16, 38% of whom went into Further Education (FE) making this the modal destination for this group. For the period under examination, 2000/01 to 2008/09, the percentage of S4 pupils staying on at school until age 18 (S6) was stable at around 45%. In the years since, this has risen rapidly and reached 63% in 2016/17 (Scottish Government, 2017b). This suggests that the profile of senior school pupils has changed, with young people from a broader spectrum of socio-economic backgrounds remaining at school, and that secondary school attainment at later levels is even more important for individuals' life chances.

Figure I.1

Scottish Education System Timeline



Between 1984 and 2013, as shown in Table I.1, the age-16 qualifications were Standard Grades (SGs). These were assessed at three different levels, Credit,[2] General and Foundation, to increase the opportunities for young people to leave school with formal qualifications at age 16. These were augmented by Intermediate (I) 1 and 2 qualifications from 2000 as part of the *Higher Still* policy reforms that aimed to increase participation in post-compulsory education (Croxford, 2009). The curriculum structure underpinning the age-16 Standard Grade and Intermediate 2 (SGI) qualifications generally comprised eight subjects: two compulsory subjects (English and Maths), five closed option choices (whereby pupils had to choose one science, humanity, modern language, technology and creative subject) and one completely free option choice that could be used to study, for

[2] Credit level grades 1 and 2 were equivalent to C and above passes at 'O' grade, the qualifications that Standard Grades replaced.

instance, a second science, humanity or modern language. This structure allowed a high degree of subject choice within a framework that effectively ensured that a minimum of four-five traditional academic subjects would be studied; that is half or just over half of an individual's S4 curriculum. In addition to the curriculum structure, subject choices at this qualification level are influenced by teacher guidance/decisions, parental input and local authority/school constraints in terms of staffing and timetabling. These latter factors increase in importance at higher levels of secondary school qualification.

Table I.1

Scottish Secondary School Qualifications

Scottish Credit & Qualifications Framework (SCQF)	
Level 7	Advanced Higher - A-C Grades
Level 6	Higher - A-C Grades
Level 5	Standard Grade (Credit 1-2), Intermediate 2 - A-C
Level 4	Standard Grade (General 3-4), Intermediate 1 - A-C
Level 3	Standard Grade (Foundation 5-6)

(Source: Scottish Government, 2017b)

Beyond compulsory education, the next level of qualification is the Scottish *Higher* (H). Highers are regarded as the 'gold bricks' of the Scottish education system. Awarded for over 100 years and reformed on a number of occasions, they have always been dual-purpose qualifications, designed to enable progression to either Higher Education (HE) or employment. They have been shown to have a distinct labour market value, attracting a wage premium, unlike 'A' Levels in England whose principal purpose is to enable HE entry (Gasteen & Houston, 2007). They are one-year qualifications taken at ages 17 and/or 18; that is in Fifth (S5) and/or Sixth Year (S6), the final two years of post-compulsory schooling. In Scotland, the Higher is the crucial qualification for HE entry. Entry requirements for more prestigious universities and/or more sought-after courses often require that five Highers at specified grades are achieved in the one sitting in S5, often in specified subjects. For instance, to gain entry to medicine in Scotland, prospective students must achieve a minimum of four A and one B grades at Higher in one sitting in S5; these must include at least two named sciences (one of which must be Chemistry), Maths and English.

Higher subject choices are made provisionally at the end of S4 and confirmed at the beginning of S5 subject to individuals' age-16 attainment, teacher guidance/judgement and school provision. There are no compulsory subjects at Higher although individuals usually take English and often Maths. Generally, pupils' S5 Higher choices are a sub-set of the subjects studied for their age-16 qualifications, and they would be allowed to proceed to Higher only if they had achieved given grades in these. For instance, with SG qualifications, it was recommended that pupils only continued to study a subject at Higher if they had achieved Credit level 1 or 2 in that subject in S4. This guidance was given on the basis of analysis of annual results' data that showed lower SG attainment to be associated with much reduced percentage pass rates at Higher. The pressures of school league tables incentivise teachers to channel pupils into the subjects that they believe individuals are most likely to pass (at all qualification levels). Not every subject offered for age-16 qualifications can be offered at Higher. Higher provision depends on both the local authority/school's ability to resource this (in terms of the availability of suitably qualified teachers and timetable constraints) and subject-specific demand.

If individuals stay on to S6, subject to approval, they can choose to take a maximum of two 'crash' Highers, that is a Higher in a subject/s that they did not study for age-16 qualification and, in principle, can take Advanced Higher (AH) qualifications if they have taken Highers in the subjects concerned. The S6 year might be regarded a 'top up' stage in terms of acquiring more qualifications for some individuals who already have what they need to progress but for others achieving additional qualifications will be a requirement for University entry. Taking Medicine as the example again, the high competition for places means that prospective students need to add appropriately to their five Highers from S5, usually requiring all three sciences at Higher and at least one science at Advanced Higher. Advanced Higher provision is, however, very constrained, even within Scotland's largest education authority area, Glasgow City Council. Where AH qualifications are required for university entry, lack of provision can impact adversely. In recent years, narrow AH provision at schools has given rise to initiatives such as the Advanced Higher Hub at Glasgow Caledonian University in the centre of the City, where pupils from schools all over Glasgow can attend to study given AH subjects not taught at their own school.

While Figure I.1 shows on-time completion of the different levels of qualification in the Scottish system, there is a high degree of flexibility of access to the study of different levels of qualification in the senior school years S5 and S6. Individuals who stay on to S5 are able to reach back to take S4 level qualifications (SGIs) if needed. This allows those who failed or did not do sufficiently well in a particular subject that they require for progression, to take this again and also, effectively, provides a second chance for individuals who performed poorly across the board to retake their S4 qualifications. Individuals in S6 can reach back to take a Higher that they require for progression but did not take while in S5 and, as discussed above, this can be a ‘crash’ Higher.

Qualifications in Scotland are housed within the Scottish Credit and Qualifications Framework (SCQF). This was superimposed on existing qualifications in 2001 with the aim of enabling objective comparison across different types of qualifications for employers and educational institutions, facilitating credit transfer for prior achievement between the latter. The SCQF has 12 levels covering all qualifications in Scotland: school, FE, HE and vocational qualifications.[3] Table I.1 above shows different SCQF levels for the qualifications taken in state secondary schools in Scotland together with their pass grades. [4] SQA is the sole national awarding body with responsibility for the accreditation of all Scottish qualifications apart from degrees. The grade boundaries for SQA qualifications are set annually at subject specific award meetings. Award panel members discuss and agree the minimum marks needed to gain a grade C in a subject and for a grade A and upper A. Grades B and D are then calculated automatically. Grade B is set at the half-way point between grades A and C, while grade D (a level of attainment close to but not a pass) is set at 80% of the grade C mark.

In 2013/14, new *National* 4 and 5 qualifications were introduced under the Curriculum for Excellence (CfE) to replace SGIs.[5] CfE was heralded as a major reform of the Scottish education system designed to ‘…. achieve a transformation in education in Scotland by

[3] See https://www.sqa.org.uk/sqa/files_ccc/OldVsNew-UpdatedJuly2013.pdf for a timeline of the introduction of SCQF and details of the qualifictions, accessed 08/08/2018.

[4] Independent (private) schools in Scotland take a mixture.e of SQA qualifications and English qualifications set by the various English examination boards.

[5] The new Curriculum for Excellence, National 4 and National 5 qualifications were introduced in the academic year 2013/14 replacing Standard Grades immediately and overlapping with Intermediates until 2014/2015 when these were last certificated.

providing a coherent, more flexible and enriched curriculum from 3 to 18.’ [6] It is supposed to provide a more holistic approach that takes account of the ‘totality of experiences’ for learners throughout their education. In practice, at the level of the new National qualifications, this implies that pupils will be entered for these when they are ready – that is earlier / later than S4. Given this major change to the system, with the introduction of new qualifications and greater flexibility in the timing of formal assessments, it is important to examine attainment immediately prior to this under the previous system to provide benchmarks against which the changes might be evaluated. This thesis helps to do this by examining the impact of socio-economic background and gender under the previous system in the years 2002-2009, just before the implementation of the CfE.

Measuring Attainment

For policy target setting and evaluation purposes, educational attainment tends to be measured in terms of a benchmark number of awards at a given level. Currently, the Scottish Government (Scottish Government, 2017a) defines the educational attainment gap as the difference between the percentage of individuals from the most deprived 20% of households and the percentage of those from the least deprived 20% achieving a given benchmark. Households are classified as such according to the Scottish Index of Multiple Deprivation (SIMD), where SIMD quintile 1 is the most deprived 20% of households and SIMD quintile 5 is the least deprived, or most affluent, 20%. The attainment gap between children from the most and least disadvantaged households increases as qualification levels rise in the SCQF from Level 5 (at the end of compulsory schooling) to Levels 6 and 7 (Highers and Advanced Highers; the qualifications required for tertiary education entry).

As can be seen from Table I.2, the percentage points’ gap between those from SIMD 1 and SIMD 5 securing one or more awards at SCQF Levels 4, 5 and 6, respectively, is 6%, 20% and 38.5% (*ibid*). It is the aim of Scottish Government educational policy to close these specific SIMD 1/SIMD 5 gaps to 2% at SCQF level 4, 5% at level 5 and 15% at level 6 by the middle of the next decade. It can also be seen that the attainment gap between males and females increases as qualification levels rise with the latter performing better across

[6] https://education.gov.scot/scottish-education-system/policy-for-scottish-education/policy-drivers/cfe-%28building-from-the-statement-appendix-incl-btc1-5%29/What%20is%20Curriculum%20for%20Excellence, last accessed 15/08/2018.

all levels. Over the two academic years shown in Table I.2, this gap is stable around 1% at SCQF Level 4, 4% at Level 5 and 11% at Level 6.

Measuring attainment in terms of a benchmark number of awards, however, is not very satisfactory for two reasons. Firstly, this implies that a given pass grade in one subject, for instance, Maths is equivalent to the same pass grade in another subject, for example Art or Physical Education. This is unsatisfactory because some subjects have more currency than others both when applying for either tertiary education courses or jobs. Secondly, it is unsatisfactory to measure relative attainment at the critical S5 and S6 level simply in terms of a pass, as the subject grades achieved are of paramount importance for university entry and/or specific course access.

Table I.2

Percentage of School Leavers by Attainment at SCQF Levels 4 to 6 and Gender / Socio-economic Background: 2014/15 and 2015/16

	2014/15			**2015/16**		
	1 or more at SCQF level 4 or better	**1 or more at SCQF level 5 or better**	**1 or more at SCQF level 6 or better**	**1 or more at SCQF level 4 or better**	**1 or more at SCQF level 5 or better**	**1 or more at SCQF level 6 or better**
Gender						
Male	95.7	83.1	54.7	95.9	83.9	56.3
Female	96.7	87.3	65.9	96.8	87.5	67.3
SIMD						
SIMD 1 [7]	92.6	74.0	41.2	92.8	74.4	42.7
SIMD 2	94.8	80.7	50.6	95.4	81.4	52.2
SIMD 3	97.1	86.4	60.3	96.8	86.7	62.2
SIMD 4	97.8	90.1	69.2	97.8	91.4	71.1
SIMD 5 [8]	98.9	94.9	80.3	98.8	94.7	81.2

(Source: Scottish Government, 2017a)

Class / Socio-economic Background

Prestigious Russell Group [9] universities advise that prospective students should take certain subjects that 'are more often required than others' for entry to their degree

[7] Most deprived 20% of households.
[8] Least deprived 20% of households.
[9] The Russell Group is a self-selected, elite association of 24 UK research universities set up in 1994 to represent their interests. In 2017, Russell Group universities received more than 75% of all university

courses (Russell Group, 2013) and hence facilitate access to these institutions. Eight subjects are identified as *Facilitating Subjects*: English, Languages, Maths, Physics, Chemistry, Biology, Geography, History. Iannelli and Klein (2015) find that subject choice is a key driver of university entry in Scotland and that working-class students take fewer academic subjects – those that *facilitate* access to HE. This suggests the existence of *secondary* inequalities in educational choices in addition to the existence of *primary* inequalities in levels of attainment. Primary inequalities arise through the relationship between social background and actual academic attainment when inequalities in family resources that are beneficial for learning cause differences in academic competencies (Stoké, 2007). Secondary inequalities occur when there are systematic differences in the educational decisions made by individuals from different social backgrounds even though they have the same level of attainment.

Education is a positional good (Breen et al., 2009); the value of qualifications at all levels will diminish as increasing numbers of individuals achieve them as evidenced by the returns to degrees in the UK after the expansion of HE. Within qualification differentiation by subject is one way of maintaining position. There is much evidence to suggest that individuals from more privileged social backgrounds tend, on average, to choose more aspirational educational options than those from less privileged backgrounds even when they have the same levels of previous attainment (Jackson, 2013a, Breen et al., 2009). If class differences in subject choices become stronger as surface level inequalities decline in the number of awards, focussing solely on the level of educational attainment will overestimate the extent to which inequalities have been addressed (Breen et al., 2009). In turn, differential subject choice can reinforce class differences in university entry and later labour market outcomes (Iannelli 2013). Social inequalities in HE entry in Scotland have been found to be explained mostly by subject choice with graduates from advantaged socio-economic backgrounds tending to choose fields of study that lead to better jobs more frequently than graduates from less advantaged backgrounds (Klein & Iannelli, 2014).

research and contract income in the UK (https://en.wikipedia.org/wiki/Russell_Group, last accessed 03/08/2019).

The transmissions mechanisms by which income inequality can impact on children's educational attainment are considered by Blanden and Gregg (2004). They make the distinction between *causal* factors (determined by income) and *non-causal* factors (associated with but not determined by income) that can result in children from low-income families achieving less at school. Adults in low income families are more likely to have the following non-causal characteristics that can put children at greater risk of low educational attainment viz:

- Lower parental education or other less easily observed adult heterogeneity (that can result in lower home-based child development)
- Poorer innate ability
- Lower emphasis on educational achievement in parenting
- Reduced ability to translate parenting time into educational development
- Shocks leading to low income such as family break-up
- Financial problems increasing family conflict/parental stress, reducing ability of parents to engage in effective parenting that improves educational outcomes

A large Behavioural Genetics literature suggests that educational attainment may be *heritable* to different degrees and, therefore, transmitted across generations. Goodman and Gregg found that almost 20% of the gap in test scores between children from the most affluent and most deprived backgrounds appeared to be explained by 'a direct link between the childhood cognitive ability of parents and that of their children.' (Goodman & Gregg, 2010, p7). If educational outcomes are determined by children's innate ability, parents' education and/or cognitive ability, parenting styles/skills and other factors that are not caused by income (but may be related to it), then, as proposed by Blanden and Gregg (2004), growing income inequality will not affect children's educational attainment. If, however, the influence of income-determined factors such as the quality of child-care, the home environment, social activities, neighbourhoods and schools is important, then growing family income inequalities will be manifested in increased inequalities in educational outcomes affecting children's life chances.

Gender

The last 50 years have seen levels of secondary school attainment in Scotland (and the UK) rising substantially for both males and females. Since the mid-1970s, however, the increase in average female attainment has been greater than that in average male attainment (Tinklin et al., 2001), resulting in a gender gap in secondary school educational performance in favour of females. [10] Focussing on male under-achievement shown by this average attainment gap, however, is too simplistic and misleading as more complex gendered patterns of subject choice, and attainment therein, lie beneath this. Whilst, in general, higher levels of female attainment are seen at each level of the school system (pre-school, primary and secondary), males and females are not two separate homogenous groups; there are high performing boys and underachieving girls. Boys' educational performance appears to be characterised by greater variance than that of girls' (Machin & Pekkarinen, 2008). Gender differences in attainment can be seen at all socio-economic levels but there is no evidence of any systematic differences between schools (Tinklin et al., 2001). Girls were found to be more likely than boys to pass almost all the subjects they studied at SCQF Levels 4, 5 and 6, and were more likely to gain 'A' grades in their Highers - the SCQF Level 6 qualifications necessary for university entry (Tinklin et al.,2001). Boys, however, were found to be more likely to achieve an 'A' grade at Higher in some subjects: modern languages, Maths, Biology Chemistry, Economics, Accounting and Finance. This appears to be in line with observed, well-established, gendered patterns of performance whereby boys tend to do better at Maths and girls at reading and writing (see, for instance, Machin & Pekkarinen, 2008). It should be noted that, in general, the impact of socio-economic background on attainment has been found to be far larger than the effect of gender.

Subject choice is starkly gendered reflecting arguably what appear to be different male and female aptitudes for Maths and language respectively. The curriculum framework underpinning age-16 qualifications ensures that Scottish pupils have a broad educational experience at this level with exposure to languages, science and technological subjects but where there is choice, this is clearly gendered. Biology has been very much the

[10] This trend is common to other developed countries (see Tinklin et al., 2001, Machin & Pekkarinen, 2008).

"female" science of choice and Physics the "male" science of choice, while technology subjects are male dominated. At Higher (SCQF Level 6), this pattern of gender choices is repeated and becomes even more stark in technology subjects. This is a long-established, deep-rooted pattern of uptake, largely unchanged in the last twenty years (see Tinklin et al., 2001 for an analysis of subject choice in 1999) and is of concern in terms of the drivers of Scottish economic growth and employment. Science, Technology, Engineering and Mathematics (STEM) employment in Scotland has been growing rapidly since 2010 and is forecast to continue to do so for the next decade (Scottish Government, 2017c). At UK level, STEM related jobs are expected to increase at twice the rate of other occupations over the next decade (UK Commission for Employment & Skills, 2016) but employers have been reporting substantial STEM recruitment difficulties and skills' shortages for some time (UK Commission for Employment & Skills, 2015). STEM jobs tend to higher paid on average (*ibid*), ergo encouraging greater numbers of females into these occupations may help to mitigate the gender pay gap whilst enhancing the pool of talent available to STEM employers. It has been estimated that doubling the number of highly qualified women in STEM employment could increase Scotland's annual national income by as much as £170m (Royal Society of Edinburgh, 2012).

Contribution and Originality

Evidence from official reports and statistics and academic studies (for instance, Scottish Government, 2017d, Sosu & Ellis, 2014, Commission on School Reform, 2013, Croxford, 2009, OECD, 2007, Raffe et al., 2006) indicates that whilst the provision of state secondary education in Scotland is equitable its impact and allocation are not. The distribution of tertiary education has increased with the expansion of the system and widening access, however, it is still the case that disproportionately fewer individuals from the most deprived 20% of households enter HE compared to the most affluent 20% of households (Commission on Widening Access, 2015). Individuals from more deprived socio-economic backgrounds systematically do less well in the Scottish education system, gaining less awards than those from more affluent backgrounds (Bradshaw, 2011, Sosu & Ellis, 2014). Of more concern, perhaps, is allocation in terms of subject choice. As outlined briefly above, there is evidence to suggest that, when given choice, children from more disadvantaged backgrounds tend to choose less academic or facilitating subjects; the subjects required for entry to more prestigious universities and courses that attract higher labour market premia (Iannelli et al., 2016). Additionally, there are clear, entrenched gendered patterns of subject choice (Scottish Government, 2017c, Tinklin et al., 2001). Early, age 14-15 decisions on subject choice can have life-long impacts for individuals' in terms of entry to tertiary education and labour market outcomes (Klein & Iannelli, 2014, Murray, 2011). Sub-optimal subject choices mean that the ability and talents of individuals are not being fully developed in the education system and the human capital of the Scottish workforce is lower than it otherwise would be.

The thesis builds on and extends previous empirical literature on educational transitions, attainment and subject choice in the Scottish education system through its original use of the full SQA administrative database for the years 2002-2009. SQA is the single awarding body in Scotland for formal secondary school qualifications (unlike England which has multiple examination boards at various qualification levels). This means that its candidate database, established in 2000, effectively provides a *census* of achievement in Scottish state secondary schools.[11] Much previous research has focussed on attainment and/or is often based on the analysis of secondary data from large-scale surveys or, more recently, administrative data linked to such surveys. Usually, narrow measures of attainment have been examined; typically, the number of awards at a given qualification level, failing to take account of the importance of both subject choice and *within* subject grades.

Through the use of national administrative data for Scotland, the thesis widens the lens on socio-economic background and gender inequalities in its state secondary schools, providing greater breadth and depth of analysis than hitherto, in the period immediately before the introduction of a major reform of the Scottish educational system in the CfE. In doing so, it provides benchmarks by which to measure the progress made in addressing inequalities under the CfE and makes several distinct and unique contributions. The application of Mare's (1981) sequential logit approach to the analysis of secondary school *academic route* transitions in Chapter One adds to the previous literature on transition analyses for Scotland as these were survey-based (e.g. Iannelli & et al., 2016, Croxford, 2009, Paterson & Iannelli, 2007, Raffe et al., 2006). A comprehensive analysis of

[11] This is not the case for independent (private) schools as their pupils sit a mixture of SQA and different English exam board qualifications.

attainment in terms of grades achieved in individual *facilitating* subjects (traditional academic subjects required for entry to more prestigious universities and/or courses) at all school qualification levels is provided in Chapter Two. Previous analyses for Scotland (e.g. Shapira and Priestly, 2018, Croxford, 2009, Croxford, 2006, Croxford & Raffe, 2007, Howieson & Iannelli, 2008, Raffe et al., 2006, Tinklin et al., 2001) have focused on the number of awards at given levels or achievement in groups of subjects (e.g Gayle et al., 2016). The investigation of a potential biological explanation for gendered subject choices in STEM subjects in Chapter Three, in terms of exposure to heightened levels of testosterone is unique, adopting an approach that has been employed previously by economists in only three other studies (Gielen & Zwiers,2018, Gielen et al., 2016, Cronqvist et al., 2016), none of which examined educational choices.

Chapter One investigates the extent of secondary inequalities in subject choice specifically in terms of the uptake of facilitating subjects at Higher (SCQF Level 6), the crucial qualifications for university entry. The aim is to ascertain whether or not, given age-16 attainment, there is secondary inequality in subject choice for formal qualifications at senior school level. In Scotland, this builds on the work of Iannelli et al. (2016) who examined subject choice using the Scottish School Leavers' Survey (SSLS) for the period 1987-2005 (and found that working class students tended to take fewer academic subjects) but uses the SQA data to investigate more widely. Mare's (1981) approach of modelling individuals' educational paths as a series of sequential decisions was adopted. Sequential logit analysis was used to model, firstly, the decision to stay on at school after age 16 and take at least one Higher and, secondly, the likelihood of choosing to study four or more Highers in Facilitating Subjects; essentially, this is the top academic track. Unlike 'A' Levels in England, the primary purpose of which lies in enabling HE entry, Highers have had a distinct labour market value in Scotland attracting a wage premium (Gasteen & Houston, 2007). Therefore, individuals' subject choices at Higher may depend on whether HE or the labour market is their intended post-school destination. These different aspirations and motivations are unobserved, so it is not possible to distinguish between them. If individuals' intended destination is the labour market then, previous attainment permitting, the facilitating Highers that are likely to be chosen would be English and Maths with other choices not being of any particular concern. If HE is the intended destination then, in general, to gain entry to more prestigious universities and/or courses, Highers in four or more Facilitating Subjects are required; accordingly, this was the subject choice threshold level that was analysed. The models included individual characteristics (gender, household SIMD, age-16 attainment in terms of total UCAS [12] Standard Grade or Intermediate 2 points gained, relative (within school) level of achievement), school characteristics (year cohort size, socio-economic composition of pupils), urban/rural location, local authority, youth employment rates and year dummy variables. Whilst age-16 attainment was found to be the main determinant of both the decision to stay on and facilitating subject choice, there was clear evidence of secondary inequalities in Higher subject choices. Given their academic achievement at age 16, individuals from the most deprived 20% of households were found to be significantly less likely to take four or more Facilitating Highers compared to those from the most affluent 20% of households.

Clearly, to progress to HE and access more prestigious universities and/or degree programmes, the *level* of attainment in facilitating subjects is important, not just the number taken. Chapter Two investigated the influence of gender and socio-economic background on attainment specifically in so-called facilitating subjects; that is, after 'good' subject choices have been made. Attainment in English, Maths, Geography, History, Modern Studies, Modern Languages, Biology, Chemistry and Physics was examined at all levels of qualification: Standard Grade/Intermediate 2, Higher and Advanced Higher. Candidates' results in the SQA data are given as grades rather than raw exam or coursework marks with the result that the subject attainment dependent variables are categorical rather than continuous. Multinomial logit models are estimated for each facilitating subject, at the different qualification levels, examining attainment in terms of the likelihood of achieving three different grades, *low*, *middle* and *high* pass, compared to a *fail*. The models included the same set of independent variables used in Chapter One apart from youth employment as this was clearly not relevant. In keeping with the general finding in other research (for example, Tinklin et al., 2001 for Scotland), the influence of socio-economic background was seen to be greater than that of gender. Good age-16 attainment was seen to increase the probability of a low pass at Higher regardless of socio-economic background but not at middle and high pass grades where those from lower

[12] Universities Central Admissions Service.

SIMDs were less likely to achieve such compared to individuals from SIMD 5. Females were seen to outperform males in most subjects across the different qualification levels but there were some notable exceptions. In Maths and named sciences at Higher, the critical level of qualification for HE entry, males were found to be significantly more likely to pass these subjects, at all pass grades rather than fail, compared to females.

In addition to persistent gendered patterns of Maths and science uptake in Scottish secondary schools, the Chapter Two results indicate that grade attainment in these subjects at Higher is also gendered. This conforms to observed patterns of differential gender performance that have come to be regarded widely as *norms* with respect to male and female aptitudes. Gender differences in educational choices and attainment have tended to be attributed to social conditioning and gender-biased environments rather than any biological predisposition. From the perspective of informing and developing educational policy, it is important to know whether such differences might be the result of *nature* as well as *nurture* because, if so, the role of socialisation and stereotyping may have been exaggerated. Chapter Three explores whether there might be a role for biology in the choice of and attainment in STEM subjects in Scottish state secondary schools. Specifically, the potential impact of variations in testosterone levels on STEM subject choices and attainment is explored by exploiting the Twin Testosterone Transfer (TTT) hypothesis that suggests that, potentially, female twins with male co-twins are exposed to increased levels of testosterone in-utero. This exogenous, random variation in testosterone levels provides the basis for a natural experiment. Twins and a control group of closely spaced sibling pairs were identified from the SQA data. Logistic and multinomial regression models were estimated to examine subject choice and attainment, respectively, in Maths, Biology, Chemistry, Physics, general Science and Computing. In the event, no evidence of any TTT effect on either STEM subject choice or attainment was found although, as expected, there were clear patterns of gender segregation in subject choice at all qualification levels. There were, however, stark social inclusion issues with respect to STEM subject choice. The likelihood of individuals from more disadvantaged backgrounds studying a named science, as opposed to general Science, together with the likelihood of achieving the better grades in these subjects, was seen to fall as household deprivation levels rose.

Administrative Data

The SQA database of candidate results was set up in 2000. A direct approach was made to SQA (in 2010) to access their candidate qualifications' database for 2000 to 2009. A formal data share agreement between SQA and Stirling University was entered into for an initial period of two years; this was reviewed subsequently and extended on two further occasions. The data are managed as three separate databases that can be linked as required, these are: candidate results, candidate details, presenting centre (school, college or other institution) details. An SQA statistician performed the data linkage work in SAS (this process took three-four full working days) and a CSV file containing the linked data was released to Stirling University in early 2010.

Investigating, cleaning and rendering the raw administrative data was a slow, painstaking process. Initially, because of their sensitive nature (containing candidate names and addresses), the data had to be held securely on a Stirling University server. The very large size of the dataset (just under 1.5 million observations containing both independent and state school and college results) meant that it could not be accessed remotely and worked on efficiently. In the event, the data were cleaned and rendered on site at Stirling University over a four-year period (2010 to 2014), part-time, one day per week. [13] Once a variable for candidate household had been derived, the data were anonymised by removing names and addresses making them portable and allowing the pace of the research to speed up. It should be noted that being allowed direct access by SQA to candidate names and addresses as an individual researcher is not possible now; for data protection reasons this information is no longer released. At the time of the second data share agreement, these details were deleted from the master data files held at Stirling as required by SQA (as indicated above, this information was not held in any of the derived datasets).

A major drawback of using administrative data for research purposes is that they are not collected for this purpose. Administrative data entry can be highly variable. Data may be entered inconsistently by different individuals across years or more superfluous data, from an administrative perspective, may not be entered at all. The result is that much

[13] The author was a full-time academic at another university until very recently.

time can be spent both rationalising and completing a dataset, as it was in this case, before derived variables can be created or the data can be added to with information from other sources. Frustratingly, these inconsistencies may often not be detected until midway through a rendering process or worse, a piece of analysis. All three component databases provided particular challenges arising from data entry inconsistencies. Table I.3 shows the data provided from the three databases and the main data challenges that were faced are outlined below.

Table I.3

SQA Database Components

SQA Databases		
Candidate Results	**Candidate Details**	**Centre Details**
Product Code (4 digit alpha/numeric code identifying the product)	Scottish Candidate Number (9 digit numeric code identifying student)	Centre Code Number (7 digit digital code)
Product Level (2 digit numeric identifying level)	Name	Centre name
Product Type (1-6 digit alpha info identifying product type) [SG / H / AH / HN]	Date of Birth [Age as at 1st Dec - in academic year can be derived]	Centre Type (Identifies colleges/schools etc)
Product Title	Address	Post Code
Result (Grade or pass etc) [Pass/Fail for HNs]	Country	
Year	Postcode	
SCQF	Gender	

Candidates' results (grades) were transformed into values for the various subjects they studied as indicated by the "Product" Codes in Table I.3. When an SQA qualification syllabus is reviewed a new Product Code is generated for the revised qualification. This meant that the same subject, for instance, Higher Economics, had multiple Product Codes that had to be rationalised before subject names could be attached for use as variable names. Once the Product Codes were rationalised and renamed as subjects, they were amended to incorporate the level of the award (e.g. Standard Grade or Higher) as indicated by the *Product type*. The different subject grades were converted into their UCAS points' equivalent to enable both comparison of attainment across qualifications and the creation of aggregate absolute and relative measures of attainment; for instance, individuals' total Standard Grade/Intermediate 2 UCAS points and school quartile in which their points' total placed them. Much of this data rendering, e.g. replacing Product Codes with subject names, was carried out using Excel spreadsheets as this enables both comprehensive data management and ease of checking.

With candidates' details, the main challenge was to create a household identifier to identify sibling groups. This was carried out by grouping individuals according to their surname, first line of address and post code using STATA's group command. The major obstacle here was the inconsistent entering of names and addresses; all of which had to checked and rationalised. [14] Gender was recoded in a zero-one dummy variable. There was no information available on race. The SIMD was merged to candidates' postcodes to use as the indicator of socio-economic background in keeping with its use in policy evaluation and formation. Initially, it had been hoped to track candidates' in terms of moving home and associated SIMD areas but only the current (last) address of a candidate was recorded, all previous addresses being replaced.

For centre details, SIMD information was merged also to schools' postcodes to indicate socio-economic environment. It was discovered, however, that where schools had been closed or merged, their postcodes were missing; this information had to be checked and entered to maximise usable observations. To analyse the decision to stay on at school after age 16 and take Highers, it was appropriate to include an indicator of youth employment by local authority area (sourced from the Labour Force Survey (LFS)). In the process of merging this information, it became apparent that schools' local authorities had not always been entered; again, to retain the maximum number of observations in the dataset, this information had to be retrieved and entered.

Scottish Index of Multiple Deprivation

The SIMD identifies and ranks small *population* areas or Data Zones in terms of their relative concentrations of deprivation across several different domains; i.e. it provides a relative measure of *multiple* deprivation. The SIMD 2009 is used in this thesis as it coincides with the last year of the data under investigation; Figure I.2 shows the

[14] For instance, the surname O'Brien could also appear variously as OBrien or Obrien. Avenue could also be entered as Ave or Av.

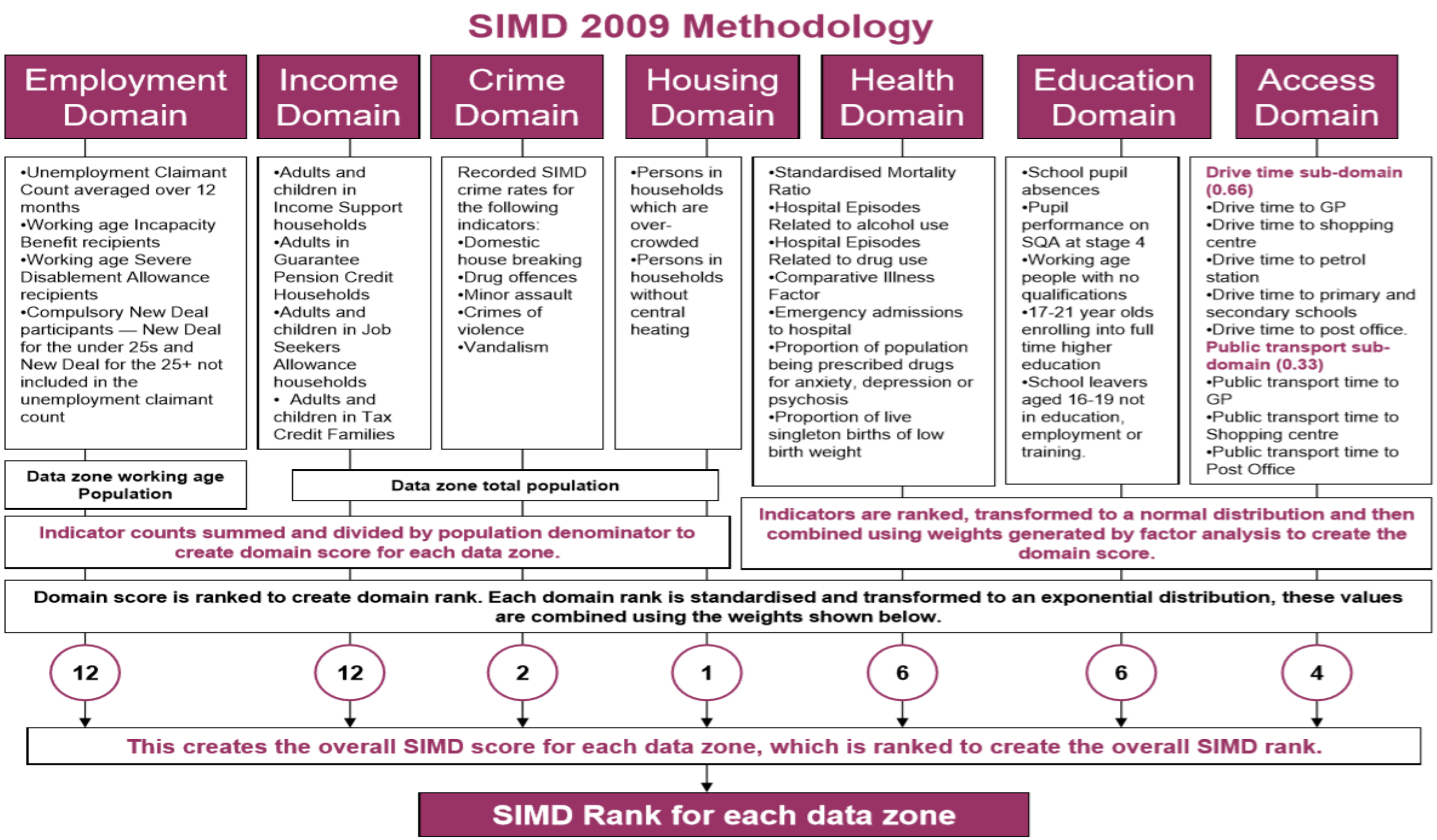
methodology employed to calculate this. Scotland has 6,505 Data Zones containing approximately 350 households and having an average population of 800 individuals. A deprivation score is calculated for each Data Zone determining its ranking from 1 (most deprived) to 6,505 (least deprived). This enables Data Zones to be compared in terms of the relative deprivation but does not indicate the extent to which one area is more deprived than another. For SIMD 2009, this score is calculated from 33 indicators across seven domains: income, employment, health, education, skills and training, housing, geographic access and crime. As indicated in Figure I.2, the overall index score is a weighted sum of the seven domain scores; domain weights are determined by their relative importance in measuring multiple deprivation, the robustness of the data and the time lag between data collection and the production of the SIMD. [15] Relative levels of deprivation tend to be reported by applying threshold cut-off points, for example by referring to the most deprived 10%, 15% or 20% of households.

SIMD quintiles are used routinely in the reporting of official Scottish Government statistics and, ergo, are adopted as an indicator of individuals' socio-economic background for the analyses in this thesis. The SIMD is not an entirely accurate indicator of children's socio-economic backgrounds as it is possible that individuals from low-income households may live in relatively affluent Data Zones and vice versa. It should be noted also that the inclusion of education in the SIMD might introduce some small degree of endogeneity in the analyses as

[15] Before weighting, the domains are standardised by ranking the scores and statistical transformation is applied to avoid high ranks and low ranks in different domains cancelling each other out. https://www.gov.scot/Topics/Statistics/SIMD/BackgroundMethodology, accessed 07/08/2018

Figure I.2

Scottish Index of Multiple Deprivation 2009 Methodology



(Source: Scottish Government [16])

[16] https://www.gov.scot/Topics/Statistics/SIMD/MethodologyVisual2009, accessed 07/-8/2018.

one of the domain indicators is pupil performance on SQA qualifications at SCQF Level 4 (Figure I.2). This is, however, one of five indicators used to calculate the domain score which is weighted at 14% (6/43 in Figure I.2) and, dependent on when the education score was calculated, will not necessarily include the age-16 achievement of all individuals in the dataset. The other education domain indicators capture potentially important influences on the aspirations and motivations of young people in terms of socio-cultural capital, family and neighbourhood effects that help to alleviate the potential for omitted variable bias (OVB). Entitlement to Free School Meals (FSM) is an alternative measure that has been used previously by researchers. It is, however, a rather coarse binary measure, providing limited insight as to the relative position of pupils (Scottish Government, 2017e) and does not capture the many relatively poor children who just miss qualifying for such (Morelli & Seaman, 2010). SIMD is a much richer variable combining information on deprivation across domains that should reflect family or neighbourhood characteristics that would be otherwise missing. Using SIMD quintiles enables the attainment within all five quintiles to be investigated as well as the gap between the most deprived and most affluent; the value of which was recognised by the Commission on Widening Access (2016).

A Note about Endogeneity

Endogeneity and/or OVB is likely to be present to some degree in all of the following analyses in terms of either missing or imperfect measures of individuals' innate ability and/or prior attainment. This is true of all research that investigates educational outcomes. Where available, primary school test scores have often been used as measures of ability or indications of prior attainment, but these are imperfect measures subject to OVB-induced endogeneity also. School effectiveness studies (and league tables), measuring average performance and/or value added, are likely to be more prone to ability-based OVB if individual characteristics are not included (Dearden, 2010). Dearden et al. (2011) found that the effectiveness of approximately 25% of English secondary schools varied significantly *across the prior ability distribution* of their pupils. In the studies that follow, where Higher and Advanced Higher outcomes are analysed, Standard Grade/Intermediate 2 performance, in terms of total UCAS points acquired, has been used as a proxy for ability/measure of prior attainment but it should be noted that, in common with the use of primary school test scores, this is likely to be subject to OVB. It is possible also that the use of SIMD might have introduced some very weak degree of endogeneity to the extent that ability may be heritable, with general cognitive ability determined in some manner by parental cognitive ability and social environment.

Thesis Structure

Chapter One explores the extent of secondary effects in Higher subject choice, using a sequential logit approach to model firstly the decision to stay on at school after age 16 to take Highers and then the choice to study four or more facilitating subjects. Chapter Two examines attainment in facilitating subjects for all levels of qualification in terms of the grades achieved using multinomial logit analysis. Chapter Three uses logistic regression to investigate gendered STEM subject choice and the potential influence of innate, biological factors, specifically testosterone levels. Each chapter provides relevant policy context, explains the methodology used when this has not been discussed earlier, outlines the model specifications and presents and analyses the results. Final conclusions and reflections are provided in the last section.

Chapter One

Primary and Secondary Educational Inequality in Scotland:

The Importance of *Subject Choice*?

1.1 Abstract

This paper chapter educational inequalities at age 16 in terms of the relationship between attainment, where *primary* inequality occurs, and subject choice, where *secondary* inequality may emerge. The aim is to ascertain whether or not, in spite of similar age-16 attainment, there is (secondary) inequality in subject choice for formal qualifications at senior school, pre-university level. The presence of substantive inequalities in entry qualifications for Higher Education and the subsequent implications for individuals' life chances means that it is important to examine subject choice in school level qualifications and not just simple, overall grade scores as many studies do. The prestigious Russell Group of UK universities maintains that traditional academic subjects tend to *facilitate* entry to their institutions. The SQA candidate database of qualifications' results from 2002-2009 is used to examine the uptake of Russell Group *Facilitating Subjects* for formal examination at ages 16 and 18 to determine the existence [and extent] of any subject choice inequalities. Where subject choice exists for age-16 qualifications, it is clear that there is a degree of secondary inequality over and above the pure attainment gap (primary inequality) as those from more deprived households are less likely to study a facilitating subject. Sequential logit analysis is used to model, firstly, the decision to stay on at school after age 16 and take *Highers* (Scottish upper secondary school qualifications that are required for university entry) and, secondly, the likelihood of choosing to study four or more Highers in Facilitating Subjects; the effective minimum level for entry to a prestigious university. The major determinant of both decisions was age-16 attainment, however, there is clear evidence of a secondary inequality effect in Higher subject choice. Those from the most deprived 20% of households were seen to be 26% less likely to take four plus Facilitating Highers when compared with those from the most affluent 20% despite having similar, good, above median academic achievement at age 16.

1.2 Introduction & Policy context

Tackling inequality in educational attainment is an enduring aim of both Scottish and UK educational policy as such inequality is embedded in the socio-economic structure. Inequality exists when educational attainment by individuals from higher income backgrounds is systematically and significantly higher relative to those from lower income backgrounds (Blanden et al., 2003). In its review of the quality and equity of education outcomes in Scotland, the OECD (2007) reported that that whilst quality of provision varies little from school to school, attainment varies widely with schools in more affluent areas achieving better levels of attainment than those in poorer areas. More recently, the Commission on School Reform (2013) reported that whilst Scottish schools delivered consistent, good performance with reasonably uniform quality of provision, the educational achievement of the most disadvantaged has not been raised. Provision is equitable but impact is not. The objective of government policy is to close this gap in educational attainment.

Over the last forty years, the context in which the Education System operates has changed enormously as outlined by Croxford (2009). Economic and industrial restructuring, the decline of manufacturing industry and the emergence of a largely service sector-based economy, has changed the nature of employment opportunities. An increasing emphasis on credentials has lead to a reduction in employment opportunities for low-attaining, minimum-age school leavers, increasing the numbers staying on after 16.[17] The full-time participation of women in the labour force has increased greatly to the extent that they now account for 37% of full-time employees and 47% of the total UK workforce. [18]

Against this changing economic and social background, various policy initiatives have been introduced in an attempt to address inequality in attainment. There have been four major curriculum and/or assessment changes. The introduction of the Standard Grade (SG) in 1984 aimed to provide appropriate awards for pupils at all levels of attainment following the Dunning Report (Scottish Education Department, 1977). The *Higher Still* reforms introduced at the end of the 1990s were designed to create a unified system of academic

[17] The age at which compulsory schooling ends in the UK and young people are able to leave full-time education.

[18] Office National Statistics (2018), UK Labour Market Statistical Bulletin, March.

and vocational qualifications to provide 'opportunity for all' (Scottish Office 1994). National Qualifications (NQs) were developed at a number of levels – Access, Intermediate, Higher, Advanced Higher – to provide appropriate courses/qualifications for all levels of ability. More recently, the Curriculum for Excellence (CfE), which replaced SG and Intermediate qualifications with new National 4 and 5 qualifications. [19]

In addition to curriculum / assessment changes, a variety of other policies have been introduced to encourage competition, parental choice, performance management and quality assurance, all with a focus on improving schools to raise attainment. More recent policies have focused on the needs of low attaining young people, particularly the "NEET" (not in employment, education or training) group. The Educational Maintenance Allowance (EMA) was introduced alongside the Higher Still Reforms to provide financial support to encourage those from lower income households to stay in full-time education after 16 to acquire more / higher level qualifications.

The latest OECD Programme for International Student Assessment (PISA) 2015 [20] results ranked Scotland's overall performance in Maths, Reading and Science as average but mean test scores fell in all three subjects (Figure 1.1). The Maths' test results have continued on a downward trajectory since PISA was established in 2000, when all three subjects were ranked above average. The 2015 results indicate a deterioration since 2012 and 2009 when Scotland's overall performance in Maths ranked as average while Reading and Science were above average. The 2012 results suggested a narrowing of the performance gap between disadvantaged and less disadvantaged pupils in all three subjects, and that the impact of disadvantage on a pupil's score in science had been reduced although there was no change in maths and reading compared to 2009. Examination of the Average Tariff score [21], a key measure of overall attainment at the end of compulsory schooling, suggests that this is not the case. Whilst the percentage increase

---

[19] The new Curriculum for Excellence, National 4 and National 5 qualifications were introduced in the academic year 2013/14 replacing Standard Grades immediately and overlapping with Intermediates until 2014/2015 when these were last certificated.

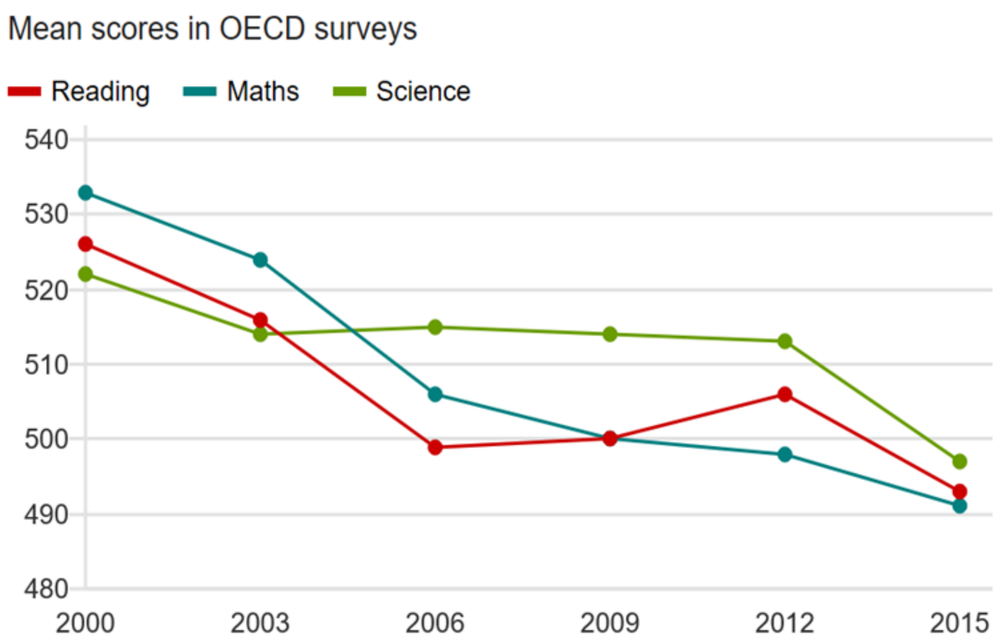
[20] The latest round of tests were conducted in March, 2015 and sat by half-a-million 15 year-olds across participating countries.

[21] The Tariff score is calculated by simply adding together all the grades (converted into tariff points) accumulated from all the different course levels and awards obtained by a student. The current tariff score scale does not recognise pupils' achievements in individual National Qualifications units and non-SQA accredited courses, and it does not include achievements of pupils in special schools.

in the Average Tariff score has been greatest for the most deprived 20% of school leavers compared with the least deprived 20%, there appears to be an entrenched absolute gap of approximately 280 points (Sosu & Ellis, 2014). The objective of government policy is to close this gap in educational attainment.

Figure 1.1

Scottish Schools' PISA Performance [22]



There is some evidence (Raffe et al., 2006) to suggest that, in line with the 2012 PISA results, the critical threshold for educational inequalities in Scotland may have shifted to age 18. The implication is that policy initiatives appear to have been successful in increasing levels of attainment and participation at age 16 across the board but have not made any significant inroads into closing the absolute attainment gap. Closer examination, however, is now required at both of these thresholds. Whilst overall levels of attainments may have increased in terms of the absolute numbers achieving the standard target of 5 awards at SCQF level 3 or above at age 16 (that is, Foundation level SG 5/6 until the introduction of Curriculum for Excellence), the subject composition of that attainment is important also. Some subjects carry more weight than others in terms of entry qualifications for HE/FE courses and are more highly regarded by employers than

[22] https://www.bbc.co.uk/news/uk-scotland-scotland-politics-38207729, accessed 13/08/18.

others. Subjects studied for formal examination/certification at 16 feed directly into subject choice for formal examination/certification at 18. To examine educational inequalities properly at these two critical thresholds, it is important to ascertain whether or not there is inequality in initial subject choice for formal qualifications' study at the age 16 threshold and whether or not this is further manifested at the age 18 threshold. As Breen et al. (2009) suggest, if there are entrenched and/or strengthening class differences in subject choice, then focussing solely on the *level* of educational achievement may overestimate the extent to which attainment has improved and/or inequalities have declined.

This study examines the extent of educational inequalities at ages 16 and 18 with respect to subject choice – secondary inequality – prior to the introduction of the CfE reforms. The SQA candidate database of qualifications' results was used which, effectively, provides a continuous census of non-degree level educational achievement in Scotland. The following section reviews the general literature and recent empirical work on subject choice for Scotland. This is followed by a discussion of the data and methodology used. The results of the analyses are then presented and discussed followed by a conclusion.

1.3 The Educational Transitions' Literature

1.3.1. Primary and Secondary Effects

The distinction between *Primary* and *Secondary Effects* in generating inequalities in educational attainment is made first in the seminal work of Boudon (1974). Primary effects operate through the relationship between social background and actual academic attainment whereby inequalities in family resources that are beneficial for learning cause differences in academic competencies (Stoké, 2007). For instance, children from more advantaged backgrounds may be subject to greater intellectual stimulation strengthening their cognitive ability, their parents may be more highly motivated and supportive of homework (Breen et al., 2009). Secondary effects manifest themselves in differences in the educational decisions made by individuals from different social backgrounds even though they have achieved the same level of attainment. Secondary effects have often been assumed to result from class differences in the costs of and returns to educational investments (Stoké, 2007). There is much evidence to suggest that individuals from more privileged social backgrounds tend, on average, to choose more ambitious educational options than those from less privileged backgrounds even when they have the same levels of previous attainment (Breen et al., 2009). Empirical research suggests that class effects on educational outcomes are split equally between primary and secondary effects (Erikson et al., 2005).

1.3.2 Educational Transitions

Informed by the insights of Boudon (1974), a large literature now exists on *Educational Transitions*, that is progression from one level of education to another. This literature is rooted in Mare's seminal modelling of educational transitions (1979, 1980, 1981). Mare (1981) emphasizes that it is imperative to distinguish between the *distribution* and *allocation* of education as changes in the former may hide changes in the latter. He argues that distribution and allocation are conceptually independent and may change over time, respectively, in response to distinct demographic and behavioural changes. For instance, education can be unequally distributed regardless of whether or not its allocation (or uptake) is random or based on gender, race or other socio-economic factors. Prior to Mare, previous work in the field variously focussed on analysing progression to different levels of formal education using linear probability models and/or the completion of different levels of education using simple linear models. Mare demonstrated that this work confounds distribution and allocation, as widening access to education was seen to disguise and/or offset changes in the relationship between education and individuals' background characteristics.

Mare (1981) argued that models of educational inequality should be specified in accordance with whether they are trying to account for changes in distribution or allocation. He maintained that educational attainment is modelled best as a sequence of discrete transitions from one stage or level of education to the next using sequential logit models to enable analysis of trends and differentials at these different stages (Mare, 2006).

> " … schooling is a sequence of events in time rather than a single status ….. sources of inequality of educational opportunity and outcome may be different at different stages of schooling." (Mare, 2006, p27)

Most importantly, given the general increase in participation and attainment noted above, this approach is invariant to changes over time in the overall distribution of education and so allows the effect of background variables on the probability of making each successive transition to be modelled (Holm & Jæger, 2011).

Mare's findings suggested that the influence of some socio-economic background variables declined over time across the transitions to different levels of education whereas previous work had indicated the impact of such to be enduring and stable. An influential, comparative, 13-country study presented in Shavit & Blossfeld (1993) reported the persistence of socioeconomic inequalities in educational attainment. In all but two of the countries, the effects of family background characteristics were seen to be persistent and stable across the different institutions and rapid expansion of education systems. The Sociology literature has continued to focus since on the empirical verification of two competing hypotheses in the context of ongoing methodological debate. Evidence of the *Waning Coefficients* hypothesis (based on the findings of Mare) and the *Constant Inequality* hypothesis (stability of impact of socio-economic variables) may be divided, to an extent, along methodological lines. A variety of different model specifications have been used since by both economists and sociologists to examine educational transitions viz: Mare sequential logit models, ordered logits, sequential and ordered probits with/without sample selection, multinomial logits and various nonparametric approaches (for a recent example of the latter, see Hu, 2017).

1.3.3 Selection Bias, Rational Choice and Alternative Approaches

For Cameron & Heckman (1998), whilst Mare's educational transitions' model represents a substantial advance on previous simple, linear models, they regard it as being only loosely motivated behaviourally with an implicit assumption of myopia on the part of economic agents at the various transition points. That is, individuals are assumed to only consider the next, immediate level of education at each transition point rather than exhibit rational choice behaviour by evaluating their medium-/long-term human capital investment based on the likelihood of expected returns. Moreover, they assert that the established finding in the transitions' literature (emanating from Mare's model), of a decline in the influence of some socio-economic background variables across the transitions to higher levels of education, is a statistical artefact that depends on the choice of statistic used to summarize the evidence (coefficients or marginal elasticities). That is, such findings are not robust to the choice of functional form. More fundamentally, the model makes no allowance for omitted variables giving rise to dynamic selection bias resulting from progressively selected samples combined with the nonlinearity of the conditional probabilities. [23]

Tam (2011) suggests that the bias introduced by unobserved heterogeneity (as identified in the Cameron & Heckman critique), would account for the pattern of declining influence of family background characteristics at higher educational transitions, suggesting reducing *inequality of educational opportunity*[24] as widely found in the Sociology literature. He suggests that the educational transitions' model is a 'convenient and useful tool for relating transition probabilities to covariates …' for descriptive purposes (Tam, 2011, pp288) but warns against its extended use to undertake comparative research across transitions, cohorts, time periods, etc..

For comparison purposes, Cameron & Heckman (1998) estimated both a standard logistic response educational transitions model and one augmented to take account of omitted variables. Little difference was found in the estimated coefficients for the first and second transitions. At higher transitions, the heterogeneity-corrected estimates were found to be greater in both absolute value and statistical significance than the uncorrected estimates, leading them to conclude that:

> "… research reporting piecemeal estimates of the schooling process tends to understate the true effects of family background on educational attainment as measured by the coefficients of logistic regression probabilities." (*ibid*, p281)

Responding to the criticisms of selection bias, whereby the logistic response model allows the effects of covariates to change freely across educational transitions when they may not vary in the population, Hauser & Andrew (2006) attempted to replicate Mare's results. They used a modified model that constrained selected social background effects to vary

---

[23] The conditional probabilities are nonlinear but can be modelled using logistic regression as the logits are linear in their parameters, assuming the variables are normally distributed. The assumption of normality, however, will be violated by progressively selected samples.

[24] Often referred to as IEO.

proportionally (with respect to the population) across transitions. The results confirmed Mare's original findings; social background was seen to explain less of the variation at each higher level of education with the influence of socioeconomic variables typically declining as students moved from one level of education to the next.

Prior to the Cameron and Heckman critique, Breen and Goldthorpe (1997) developed a formal mathematical model incorporating rational behavioural choice, vis-à-vis the expected benefits and costs of acquiring post-compulsory education, to explain observed empirical trends. In particular, they aimed to account for the observed pattern of class differentials against a background of educational expansion, increasing participation rates and the rapid decline of gender differentials in attainment. They incorporate secondary effects: the actual choices made in the educational system, maintaining that these are an important source of class differentials over and above primary effects. These choices are assumed to be the outcome of rational decision-making based on student/parent evaluation of the benefits and costs of possible alternatives. They suggest that there are three ways in which class differentials in educational attainment may occur through secondary effects. In addition to differences in resources and differences in ability and expectations of success, they also emphasize the importance of relative risk aversion in terms of maintaining the socio-economic status of the family and *anticipated* educational choices (Erikson & Jonsson, 1996). Compared to working-class children, middle-class children, it is argued, are likely to have stronger preferences to stay on at school as their families have more to lose in socio-economic terms if they leave and do not progress to the next educational level/s and associated occupations. Breen and Goldthorpe (1997) argue, however, that persisting class differentials are, in the main, explained by persisting inequalities in class resources.

To test the three Breen-Goldthorpe Model mechanisms for generating secondary effects' class differences in attainment, Stoké (2007) uses logistic regression with panel data for Germany to analyse secondary school choices in an educational system where parents have a high degree of freedom of choice over this. The findings revealed that: perceptions of the financial burden of continuing education were entirely the result of differences in family income and the number of siblings, the subjective probability of successful degree completion increased strongly with social class,[25] parents cared about family status' maintenance but this motive was not consistently strong across all classes, differences between degree choice in terms of incentives to select more ambitious educational routes grew with higher class positions. With respect to the hypothesis of rational choice of secondary school/educational route being determined by probability of success, risk aversion/social status maintenance and costs, both the probability of success and the motive of status maintenance were seen to have strong, significant effects on school choice with the former having the larger impact. Parents' anticipated costs of educational investments were seen to be irrelevant suggesting, Stoké asserts, that anticipated higher costs are unlikely to be the mechanism whereby insufficient financial resources and higher numbers of siblings influence educational outcomes as in some previous research. Inequality in educational outcomes, he further asserts, is not solely the result of rational decision making (Stoké, 2007).

Chevalier & Lanot (2002), use Cameron & Heckman's (1998) ordered probit model to distinguish between the direct and indirect effect of family income on educational decisions. Data for two different cohorts of children, born in 1958 and 1970 from the National Child Development Survey (NCDS) and British Cohort Survey respectively, are used to test the stability of results. Young people from poorer families were seen to be less likely to invest in education. Simulating a financial support policy, they find that an education benefit (to increase family education budgets) would not lead to an increase in schooling and conclude, similar to Stoké (2007), that family characteristics dominate financial constraint effects.

Multinomial analysis is used by Breen & Jonsson (2000) to examine parallel academic and vocational pathways in the Swedish school system and the probability of transferring between these at higher levels. They posit that the Mare model is unsuited to any analysis of the parallel pathways common in European school systems as it assumes exclusive, single route progression through the educational system. They examine *path dependence* – the degree to which transition probabilities between the levels and types of education

[25] Where these differences were almost completely explained by children's grade points and the type of secondary school (academic or vocational) that was recommended for them at the end of primary school (Stoke, 2007).

are contingent on students' previous pathways– and variations in the influence of class origins on the probability of choosing different educational options. Transition probabilities to the different levels/types of education were seen to be influenced by both class origins and previously followed educational pathways. Social background inequalities were found to differ among educational transitions and appeared to be stronger for those transitions where the risk of social demotion was higher. Results from the multinomial transition model were compared to those from a standard, stay-leave Mare model. The Mare model was seen to underestimate class origin influences at the first and second transitions (in keeping with other empirical evaluation, e.g. Cameron & Heckman 1998) but to overestimate them at the HE transition.

Karlson (2011) adopts the same approach for the Danish educational system which also has academic and vocational secondary school pathways. The influence of family background and individual characteristics on pathway choice are investigated using a multinomial transition model that explicitly addresses unobserved heterogeneity by including instrumental variables at the appropriate transition points (in the secondary education pathway choice). The instrumental variables, chosen as indicators of peer influence on educational decisions, are the percentages of an individual's school class that chose either the academic or the vocational track. Comparison of results from Karlson's multinomial transition model with unobserved heterogeneity with a Breen and Jonsson (2000) specification for the Danish data revealed marked social selection for both the secondary school academic track and HE transitions. The Breen and Jonsson model was found to underestimate the estimates on family background and individual characteristics. Karlson speculates that the different findings may reflect the difference between testing results for robustness to unobserved heterogeneity as opposed to explicitly accounting for it.

Ordered logit models are used by Breen et al (2009) in an attempt to re-examine the evidence for persistent inequality posited in Shavit and Blossfeld (1993). They suggest that, given the post-World War Two education and welfare policies, declining (rather than stable) disparities in educational attainment between the advantaged and disadvantaged would be expected. The research brought together data sets from eight European countries [26] for five birth cohorts born in the first two thirds of the 20th century, the last cohort being 1955-1964 and therefore turning 16 in 1971-1980. A common educational categorization and, by and large, common definitions of class were used. The joint distributions of class origins and highest level of education achieved were modelled using an ordered logit as the available data were unsuitable to estimate either a Mare model or Breen & Jonsson's (2000) extended educational transition model.[27] Social class advantages in children's educational careers were seen to be less pronounced with the decline in inequality largely occurring in the middle of the century. Breen et al. point out, however, that a declining association between class and educational attainment may not imply declining class inequality if there are '… distinctions within … educational categories that are consequential for life chances.' (Breen et al., 2009, p1515). They assert that differences between classes in choice of subjects of study or field of education have often been found and should such differences increase as inequalities in the level of education decline, then focussing solely on levels of education reached will overestimate the extent to which educational inequalities have been reduced.

The influence of social background on the quality or type of education received at different levels of the education system is the concern of Lucas (2001) who draws together two hitherto separate US educational research strands: educational transitions and *tracking*. In a British context, tracking may be thought of as akin to 'streaming' or 'setting' whereby pupils of differing levels of ability follow different educational routes in terms of subjects and/or levels. [28] Tracking in the US is the result of separate, yearly, subject-specific decisions and, therefore, Lucas argues, educational transition models need to be extended to account for type of education pursued as "… students who decide to continue also decide within which curriculum they will continue." (Lucas, 2001, p1651). Lucas maintains that, given the nature of tracking, synthesis of the educational transitions and tracking research in this manner supports the sequential decision-making model, invalidating the assertion of Cameron and Heckman (1998) that the model requires a

[26] Britain, France, Germany, Italy, Ireland, the Netherlands, Poland, Sweden.
[27] The lack of a unified sequential ordering of educational categories excluded use of the Mare model. The absence of information on individuals' educational pathways meant that the extended educational transition model of Breen & Jonsson (2000) could not be applied.
[28] A Scottish example would be where less able pupils might be encouraged to take general Science instead of a named science at the previous Standard Grade or Intermediate 2 levels of qualification.

behavioural assumption of myopia. Ordered probit models were estimated, using US High School and Beyond data for 1980, 1982 and 1986, to investigate both school continuation and *track* mobility in terms of the assignment of students to different subjects and levels. Lucas finds significant social background effects for each year studied. Social background effects were found to determine, firstly, who completes a level of education if completion of the level is not universal and, secondly, the kind of education individuals received within near-universal levels of education.

Using the 1970 BCS and the British Household Panel Survey to compare the educational attainment of 16-year olds in 1986 with those in the mid- to late 1990s, Blanden and Gregg (2004) employ a variety of approaches to isolate the impact of family income on such. Ordered probit models of qualifications that control for individual and family characteristics, a sibling fixed-effects' model and other specifications with identification strategies based on transitory income variations within the family, are used to provide a range of estimates showing the impact on GCSE attainment, staying on at school beyond age 16 and degree attainment. The results provide consistent evidence that family income impacts significantly on educational attainment in the UK. A one-third reduction in income from the mean was seen to increase the probability of a child obtaining *no* GCSE passes by approximately three to four percent on average and decrease the probability of a child staying on at school by the same amount. Further analysis, comparing educational outcomes for young people at the 10th and 90th percentiles of the income distribution, suggests that the greater the existing level of inequality, the stronger are the effects of changes in the relationship between education and income. Their ordered probit analysis suggests that at the same time educational opportunities widened at lower qualification levels, the income-education relationship strengthened at HE level. This is consistent with the findings of Blanden et al. (2003) and Blanden et al. (2004) who found a closing of the inequality gap in staying on after age 16 between rich and poor families but no similar fall in HE inequalities.

The transition to A-level courses in schools in England and Wales is examined by Jackson et al (2007) at three time-points using logistic regression analysis with data from the NCDS for 1974 and the Youth Cohort Studies (YCS) for 1986 and 2001. They created comparable variables for class background, prior academic attainment and transition choice at age 16 in terms of whether or not to stay on at school and take A-levels. Prior academic attainment was measured by O-level, CSE and GCSE English and Maths' grades only as these subjects have almost uniform uptake and this avoids the problem of averaging grades over widely differing numbers and combinations of subjects. Initial analysis identified both primary and secondary effects. Expected primary effects are in evidence as students from more privileged backgrounds were seen to have higher levels of age-16 academic attainment. Clear secondary effects were seen at intermediate levels of attainment, where those from the most privileged backgrounds were some 15-20 percent more likely to make the transition to A-level work than those from the least privileged backgrounds. A subsequent Oaxaca-type decomposition analysis suggested that secondary effects might be responsible for between 25% to 50% of observed class background differentials in A-level transition propensities. The authors suggest that it is a serious error to focus solely on differences in academic attainment as, over and above these differences, further social background differences occur in the educational choices made by individuals. Individuals from disadvantaged backgrounds tend to be less inclined to pursue more academic educational routes even when they have good previous attainment that would allow them to do so (Jackson et al, 2007).

Post-16 educational transitions in England are considered also by Moulton et al. (2018) who examined the influence of different 14-16 curriculum pathways on variously: progression to full-time education after age 16, enrolment for A-levels in general and the study of two or more facilitating A-levels. They used data from the 'Next Steps' study of 16,000 individuals born in England between 1989 and 1990 that were linked to the national administrative education data (the National Pupil Database, NPD) - to estimate a number of logistic regression models for each transition. Separate models for each outcome were estimated in stages, sequentially adding different explanatory variables (individual/family characteristics, 14-16 curriculum features, school characteristics, Key stage 4 attainment) marginal effects at means reported for comparison purposes. It was found that those individuals who had studied an English Baccalaureate [29] eligible curriculum (comprised of GCSEs in English, Mathematics, History/Geography, two sciences, one language) had a greater chance of progressing to any of the post-16

[29] Also know as an EBacc.

outcomes compared to those who had taken at least one applied GCSE subject applied (Art and Design, Applied Business, Engineering, Health and Social Care, Applied ICT, Leisure and Tourism, Manufacturing, Applied Science). An EBacc curriculum was seen to increase the probability of studying for A-levels/two or more facilitating A-levels by 29% and 20% respectively, while studying one or more applied subjects reduced the chances of such by 16% and 13%. Social class differences in progression were not explained by curriculum differences. While working class individuals were less likely to study EBacc subjects and less likely to remain in education post-16, pursuing an academic curriculum was found to be equally valuable across social classes. They suggest cautiously, therefore that encouraging individuals from working class backgrounds to take more academic subjects might act as a lever to increase their post-16 educational participation but warn at the same time that, if following an EBacc curriculum is a signal of ability, increased uptake of such may simply weaken this signal.

The role that schools play in shaping the age 14-16 subject choices of individuals is explored by Anders et al. (2018). They investigate the extent to which schools' subject provision is influenced by their pupil composition with respect to academic attainment, socio-economic background and gender. Multi-level models are estimated with NDP administrative data for the academic year 2005/06 to decompose the variance in outcomes resulting from prior attainment, socio-economic status (SES) and gender for a continuous measure of *academic subject selectivity* indicating the propensity of individuals to choose more academic subjects. Binary logistic regression models are estimated to investigate the likelihood of individuals variously studying for 3-5 facilitating subjects, a full set of EBacc subjects and lastly, all three named sciences. The subjects that individuals studied at GCSE level were found to be associated with their prior attainment, socio-economic background and gender. In addition, it was found that individuals who attended schools with higher proportions of socio-economically advantaged pupils were more likely to study academic subjects and that this SES school composition effect was similar in strength to an individual's family SES effect. Overall, schools were found to account for approximately one third of the variation in the academic subject selectivity measure, falling to a quarter once school demographics were accounted for. The authors speculate as to two potential explanations for their findings: firstly, that schools may try to tailor the curriculum they offer to the socio-economic composition of their pupils, reflecting what they consider to be appropriate and, secondly, that recruitment and retention problems faced by schools in more disadvantaged communities may constrain curriculum offerings particularly in the shortage areas of science and languages. A negative association between being in a local authority area where significant numbers of pupils attended selective schools and the academic selectivity of subjects was found, leading them to suggest that further expansion of selective education may increase horizontal inequalities in subject choices.

Both simulated and NCDS data are used by Holm and Jǣger (2011) to compare sequential probits (rather than logits) in the Mare tradition with sequential probit selection models. They examined two educational transitions: the first from age-16 into A-levels, the second from A-levels into HE. Frequencies from the NCDS data (covering individuals who would have been 16 in 1974) suggested that the 'First' transition from post compulsory study into A-levels was highly selective with just 38% of respondents completing these, whilst the 'Second' transition, from A-levels into HE is not with more than 80% of those completing A-levels moving on to HE. The simulations demonstrated that the Mare model produces increasingly biased results with each subsequent progression to a higher transition (as the strength of selection on unobserved variables increases). The probit sample selection model was found to produce consistent results with the simulated data with selection on unobservables. Its ability to do so with real data, however, was seen to be highly dependent on data quality and the amount of information available in the data. The estimated 'naïve' Mare probit models were found to suggest that the influence of family background decreased across the two educational transitions examined in keeping with the waning coefficients' hypothesis. The final probit sample selection model that included, cognitive ability, transition-specific instrumental variables, indicated the influence of family background to be constant, for the most part, across educational transitions in keeping with the constant inequality hypothesis.

Success and failure in the timely completion of compulsory education in Spain and the respective, subsequent transitions to either post-compulsory education or repeating the level are explored by Bernardi (2012). It is hypothesized that for students who fail to complete compulsory education on time, at age-16, there may be a *compensatory* effect of social background in terms of the likelihood of having a *second chance* to remain at school to retake the level. [30] Bernardi emphasizes that whilst observed differences in socio-economic inequality across educational transitions may be the result of a family background compensatory effect, they may result also from selection bias from two principal unobserved sources: students' cognitive and non-cognitive skills (Cameron & Heckman, 1998) and their anticipated choices of dropping out of the education system (Erikson & Jonsson, 1996). Sequential probits with and without sample selection were used to test the proposition that individuals' social backgrounds may *compensate* for initial failure to complete compulsory education on time (by affording them a second chance) and to address potential selection bias.

Consistent with the suggestion that unobserved ability might explaining the waning social class effect at later educational transitions (Cameron & Heckman, 1998, Holm & Jǣger, 2011), for those who completed compulsory education on time, social class effects were seen to increase when controlling for selection. On the other hand, the social background effect on the compensatory transition of staying on to complete compulsory education was seen to decrease after controlling for selection bias. Bernardi argues that this may be the result of unobserved anticipated choices that drive both the initial failure to complete compulsory education and the subsequent compensatory transition decision of whether or not to repeat. Ignoring this potential bias runs the risk that estimates of primary effects are likely to be inflated at the expense of secondary effects since the former are, in part, influenced by the latter to the extent that these are anticipated.

The influence of family background on remaining at school to complete compulsory education was found to be stronger than its effect on subsequent 'on-time' progression to post compulsory education, lending support to the hypothesis that individuals from more advantaged backgrounds would have a greater 'second chance' likelihood. This leads Bernardi to suggest that "…. a large amount of the observed class inequality in educational opportunities would come about not among 'good students' but among 'bad or not very good' ones." (Bernardi, 2012, p170). This is consistent with the findings of

[30] In the Spanish education system, students are required to pass three subjects at the end of each year of compulsory education to progress to the next level. Generally, failure would mean having to retake the subjects the following year and, therefore, would result in an individual not having fully completed compulsory education at age 16.

Breen & Jonsson (2000) that social background effects will be stronger for transitions that involve a higher risk of social demotion and accords with the views of Lucas that

> " …. If social background can move an otherwise "average" student over a threshold, then social background effectively maintains inequality." (Lucas, 2001, p1680).

Social (or *cultural*) capital has a crucial influence on educational choices (secondary effects) over and above its effect on educational performance and ability (primary effects). Middle-class families, it has been argued (Perry & Francis, 2010), have a better understanding of the way the educational system works, having acquired more social capital through educational experience and well-resourced networks. Socio-economically advantaged parents are more likely to recognise the pivotal educational decision points from their own experiences. Less-advantaged parents may be equally as encouraging and supportive of their children but they may not realise the significance of certain decisions, for instance, the importance of subject choice in senior school, particularly if they have not been to university.

1.3.4 Empirical work for Scotland

In terms of primary inequality in Scotland, whilst achievement has risen in general, a stubborn, poverty-related attainment gap persists as discussed in more detail in Chapter Two (2.2). Policy initiatives appear to have been successful in increasing levels of attainment and participation at age 16 across the board but have not made any significant inroads into closing the absolute attainment gap between the most deprived 20% of households and the least deprived 20%. Comparing age 18 performance with age 16 performance has shown the class gap to be wider at 18 than at 16 from the mid-1980s to the mid-2000s (Raffe et al., 2006, Croxford, 2009). Controlling for educational attainment, Gayle et al. (2002) found significant differences in HE participation between different socio-economic groups in the UK, suggesting that the choices individuals make at age 18 play a part in explaining differential participation. Educational choices at 18 may be limited by earlier, secondary school subject choices. Recent analysis of the Scottish School Leavers' Survey (SSLS) data (Iannelli et al., 2016, outlined below) has shown that working class students take fewer academic subjects – those that *facilitate* access to Higher Education - suggesting the existence of secondary educational inequalities. This is likely to be a significant factor in the HE entry inequalities identified by Raffe (2000). The potential exists for such inequalities to widen with the post-CfE narrowing of the S4 curriculum identified by Scott (2015) and Shapira and Priestly (2018) and, simultaneously, given the English experience (Sullivan et al., 2010), increasingly varied, individually *chosen*, tailored curriculum paths. Examining publicly available SQA data [31], Shapira and Priestly (2018) found that the average number of subjects taken by S4 students for formal (National 5) examinations was 3.7 in 2016 compared to 5.8 in 2013 prior to the introduction of CfE. There was no uniform provision of the number of subjects for National 5 study across LAs. Schools in the most deprived LAs (as measured by SIMD) were found to offer between 5-8 or 6-8 subjects for National 5 study, while the LAs whose schools uniformly offered eight subjects were in the least deprived 20%. Sullivan et al. (2010) argue that changes in the English education system, resulting from the discourse of increased choice and personal learning, have exacerbated *horizonal inequalities* in the subjects studied by individuals who are '… now directed into varied curriculum paths from a young age' (Sullivan et al, 2010., p18). This, they assert, ignores the social structures within which choices are made. Iannelli (2013) argues that student flexibility in choosing the type and number of subjects in upper secondary education may lead to substantive differences among students from different social classes in the choice of school subjects. In turn, this can reinforce class-of-origin differences in HE entry and later labour market outcomes.

SSLS and the Irish School Leavers Survey data for 1987-2005, were used by Iannelli et al. (2016) to attempt to measure the extent to which upper secondary subject choice explains class-of-origin differences in HE entry and access to different HE institutions. Logistic and multinomial regression analyses were used to examine class differences in the chances of entering HE and the chances of attending different HE institutions. The importance of subject choice in reproducing such varied between the two countries. Working-class Scottish pupils consistently take fewer facilitating academic subjects than their Irish counterparts. The average number of facilitating subjects taken by pupils from a

[31] Publicly available SQA data provides: the number of entries to SCQF levels 3-7 by individual subject, the number of pupils entered by subject at the different SCQF levels, the number of A*-C passes by subject at the different SCQF levels. The Scottish Government uses SQA data to report on school leaver attainment at LA level. Shapira & Priestly used this rendered data together with the directly available SQA data to create their time series.

professional background was four compared to two-three for those from a working-class background and this was seen to be stable over the period. The level of social inequality in HE entry does appear to have fallen over time but the reduction is less in Scotland than Ireland. Students from upper-middle class backgrounds were almost 30% more likely to enter HE at end of 1980s than those from working class background in both countries (29% Scotland, 28% Ireland). By 2005, upper-middle class Scottish students were 21% more likely to enter HE compared with 15% in Ireland. Social inequalities in Scottish HE entry were seen to be explained mostly by subject choice whereas these were associated strongly with academic performance in Ireland. Subject choice was found to explain 57%-70% of the gross association between socio-economic background and HE entry in Scotland. They concluded that 'subject choices are the main mechanism by which family (dis)advantage in the chances of entering HE (in particular, in entering the most prestigious institutions) is transmitted' in Scotland (Iannelli et al., 2016, p576). They assert that the provision of clear subject choice advice to pupils from less advantaged socio-economic backgrounds has the potential to mitigate inequalities in HE entry. In addition, when examining HE entry by institution type, it was found that while the distribution of HE had increased (though widening access and conferring university status on the former polytechnics), allocation, in terms of access to the more prestigious institutions, was reduced with social inequalities in entry to the ancient universities rising from 15 to 22 percentage points.

The above work demonstrates clearly that it is not just overall attainment grades/scores that are important, subject choice matters. Subject specific grades are required for entry into certain fields of study. Early decisions on subject choice impact on HE entry. Additionally, graduates from advantaged socio-economic backgrounds choose fields of study that lead to better jobs more frequently than graduates from less advantaged backgrounds (Klein & Iannelli, 2014). The Russell Group of UK universities advises that prospective students should take certain subjects that 'are more often required than others' for entry to degree courses (Russell Group, 2013) and hence facilitate access to their institutions. Eight subjects are identified as *Facilitating Subjects*: English, Languages, Maths, Physics, Chemistry, Biology, Geography, History. The work of Iannelli et al. (2016) indicates that subject choice is a key driver of HE entry in Scotland. If class differences in subject choice become stronger as the distribution of education increases, then, as Breen et al. (2009) suggest, a focus solely on the level of attainment will overestimate the extent to which inequalities have declined. Below, the Mare sequential logit approach is used to examine the transition to upper secondary school qualifications and the subject choice of such in state schools in Scotland using SQA administrative data for the period 2002-2009.

1.4 Data and Methodology

As SQA is the single awarding body for state schools in Scotland, its administrative data effectively provide a census of formal qualifications' choice and attainment in state secondary schools. Data from 2002 to 2009 are used to examine the uptake of Russell Group Facilitating Subjects: English, Maths, Physics, Chemistry, Biology, History, Geography, Modern Language/s. [32] The eight years of data provide 1,021,470 observations on 489,468 individuals with: 166,136 observed once (S4), 114,662 observed twice (S5), 208,670 observed three times (S6). [33] Firstly, descriptive statistics are used to establish the baseline pattern of subject choice for formal examinations taken at age 16 (the end of compulsory schooling) by SIMD quintiles. An age 17-18 snapshot of the uptake of Facilitating Subjects for formal examination at Higher (the qualifications necessary for university entry) across SIMD quintiles is also provided. Then the sequential logit approach is adopted to model the two immediate post-16 educational decisions: whether or not to stay on at school to progress to the next qualification level - Highers and, if so, what to study in terms of the number of facilitating Highers.

1.4.1 Logistic Regression and the Modelling of Choice

Standard regression analysis (ordinary least squares - OLS) assumes a continuous dependent variable (Y) that is a linear function of the independent variables (X) included in a model, and that the dependent variable and the error term (u) are normally distributed, viz:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + e$$

or: $Y = BX + e$

[32] Although the database contains data for the ten years 2000 to 2009, since its inception, the first two years have not been used as the migration to the system appears to have been incomplete with the result that these years do not provide comparable numbers of observations with the latter years.
[33] Appendix A, Table A1.

When modelling choice(s), standard regression analysis is not appropriate as the dependent variable is categorical rather than continuous and, therefore, neither it nor the error are normally distributed. For instance, the decision to stay on at school is a discrete choice that may be expressed by the dependent variable (Y) taking a value of one if an individual chooses to remain at school or zero if they decide to leave:

$Y_i = 1$ (individual *i* chooses to stay on at school after age 16)

$Y_i = 0$ (individual *i* chooses to leave school at age 16)

Modelling this choice using OLS would give a linear probability model that assumes the probability of an individual staying on (Y=1) would change linearly with the values of the independent variables regardless of how small or large these are. With a discrete choice between two alternatives this most certainly will not be the case: a relatively small increase in an independent variable (X_i) may tip the balance in favour of Y=1 (an individual staying on) such that this becomes almost certain, rising exponentially after a certain threshold level. This means that as the value of an independent variable becomes very large, Y will approach a value of one at a slower and slower rate; or as the value of an independent variable becomes very small, Y will approach zero at a slower and slower rate (Gujarati, 2015). The modelling of choice(s) using OLS, therefore, is inappropriate as the coefficients for the independent variables are not linear.

A discrete choice between two alternatives giving a binary dependent variable is modelled far more appropriately by logistic regression. Logistic regression (logit) models relate the log odds of a binary outcome measure to the explanatory variables used. The dependent variable is always a 0-1 outcome while the independent variables can be categorical or continuous. Coefficients give the *change in the log odds* or *percentage change in the odds* for a *successful* outcome (ie that the dependent variable equals 1) for a one unit change in an independent variable (Mare, 1981).

Odds ratios relate the probability of one outcome (Y=1) to the probability of an alternative outcome (Y=0). If *p* is the probability that Y equals one, then the probability that y equals zero is given by 1- *p* and the odds ratio will be:

$$\frac{p}{(1-p)}$$

The above expression implies that, if the probability that Y=1 (an individual stays on at school) is 50%, then the probability that Y=0 (the individual will leave) will be 50% also and the odds ratio will be equal to one; this would imply that either outcome is equally likely. If the odds ratio is greater than one, then Y=1 (the individual staying on at school) is the more likely outcome. Alternatively, if the odds ratio is less than one, then Y=0 (the individual will leave) is the more likely outcome. The distribution of the dependent variable (Y) determines *p*, with a cumulative distribution function for Y indicating the points at which one outcome is more likely than another. Logit models assume that the underlying probability distribution is the logistic probability distribution which gives:

$$p = \frac{1}{1+e^{-Z_i}}$$

Where $Z_i = BX_i + e_i$, that is, the vector of independent variables that determine Y.

In turn, assuming symmetry, the odds ratio can be expressed as:

$$\frac{p}{(1-p)} = \left(\frac{\left(\frac{1}{1+e^{-Z^i}}\right)}{\left(\frac{1}{1+e^{Z^i}}\right)} \right) = e^{Z_i}$$

Taking the log of this gives the logistic regression model form:

$$L_i = ln\left(\frac{p_i}{1-p_i}\right) = Z_i = BX_i + e_i$$

or: $$ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots \beta_k x_{ki} + e_i$$

Where p_i is the probability associated with the binary outcome measure for individual *i* and $e^{\beta 1}$ is the odds ratio for the effect of the independent variable x_1 on that outcome. The odds ratio ($e^{\beta 1}$) being the change in the odds of a successful outcome (e.g. Y=1, staying on at school) associated with a unit change in x_1. This transformation gives an odds ratio that is linear in both the parameters and the independent variables, allowing estimation of the outcome probability as a continuous variable (Gujarati, 2015).

Modelling post-compulsory schooling as a series of separate, sequential decisions (or *transitions* in keeping with Mare, 1981) enables the analysis of trends and differentials at various decision points. This sequential approach is made possible because the continuation probabilities are asymptotically independent of one another (Hauser & Andrew, 2006). Such specifications have been seen to deliver improved model fits over all-encompassing multinomial decision models. [34] Moreover, it has been suggested that multinomial logit models are of limited value in analysing educational transitions because by specifying multiple, potentially ordered, categorical outcomes, the conditional risk of the transitions is not captured (Hauser & Andrew, 2006).

Crucially, the specific use of logistic response models, allows the impact of socio-economic variables to be identified despite the changing marginal distributions of either the independent or dependent variables (Mare, 1981). This is extremely important as changes in the distribution of education, for instance the expansion of Scottish and UK post-compulsory secondary and Further/Higher Education, may mask changes in the influence of socio-economic background characteristics and even suggest that their influence has declined as student numbers increase. This necessitates the use of concepts or measures that can identify stability or change in inequality even when there are

[34] http://data.princeton.edu/wws509/stata/c6s4.html, last accessed 15/08/2018.

changes in the social class composition of the age group or in the total proportions achieving given education levels (Raffe et al., 2006).

As noted above, this approach has been criticised principally by Cameron and Heckman (1998) on the grounds that nonlinear conditional probabilities combined with progressively selected samples may suffer from OVB. Their comparison of a standard logistic response model with one corrected for heterogeneity, however, revealed little difference in the coefficients of the models for the first and second transitions. This suggests that, given that only two, sequentially close, decisions are being examined in this paper, the approach is still appropriate. The initial distribution of unobservable variables is independent of the explanatory variables that are included; that is, the problem only starts to emerge at the second transition. Moreover, the Cameron and Heckman comparative analysis suggests that uncorrected models will tend to understate the family background effects on educational attainment; coefficients will exhibit *downward* bias. The influence of individuals' background characteristics in the second decision/transition model (the decision to study for four or more Facilitating Highers) therefore, may be understated but not significantly so. That said, Holm and Jæger (2011) emphasize that bias arising from increasingly selected samples at higher educational transitions mean that it is important to model *entire* educational careers and not just the later transitions that may be of primary interest. Arguably, the analysis here is guilty of not modelling two earlier transitions: pre-school to primary school, primary school to secondary school. Not only is this not possible with a data set that solely records attainment of national qualifications at state secondary schools but these earlier transitions are compulsory rather than selective as a result of individual decisions. Therefore, the first choice-based education transition decision is that of compulsory to post-compulsory schooling (or not).

Tam (2011), whilst also critical of the sequential logit approach for the same reasons as Cameron and Heckman (1998), views it as a flexible, convenient tool for descriptive purposes in terms of relating transition probabilities to covariates. It is posited that this is its purpose here as the first logit models the decision to stay on to take at least one Higher (the first transition) but the second logit, in modelling the uptake of four or more facilitating subjects, is effectively *describing* the nature of the first decision (transition); that is *how* the transition is completed in terms of subject choice. For Breen et al. (2009), the first transition is that of post compulsory school choice, while the second transition is secondary completion. It should be noted that the administrative data used here only show *completions* in terms of examination entry and subsequent results, they do not reveal student drop-outs.

The use of probit selection models (with standard probits for comparison) has come to be preferred in the more recent literature as such models can generate a summary measure of the importance of selection on unobserved variables, ρ. Estimation of ρ makes it possible to account for the likelihood that unobserved variables that influence the propensity to make the first transition may be correlated with unobserved variables that influence the propensity to make the second transition (Holm & Jæger, 2011). It is not possible to estimate ρ for logit specifications as there is no bivariate logistic distribution.

Holm and Jæger (2011) also point out that the estimated effects of explanatory variables at the second transition also suffer bias as a result of scaling effects that produce *upward* bias. In logit/probit models, it is the regression coefficients divided by the error variance that are identified. [35] Scaling effects occur as the error variance in the selected sample (i.e. in subsequent transitions) will be lower than the error variance in the whole sample as the selected individuals become more homogeneous, causing upward bias. [36] Mare models are subject to *attenuation bias* caused by selection effects on the one hand, and scaling effects on the other. Probit selection models accommodate selection bias by imposing functional form assumptions on the selection equation. They do not address scaling effects however, and these cannot be separated from selection effects. In many of the comparative approaches educational transitions' literature, there appears to be demonstrable downward bias in Mare-type models suggesting that selection effects dominate scaling effects. Again, this would imply that the impact of background characteristics in the second decision model may be understated but that any aggregate bias should be relatively low.

[35] As opposed to the actual coefficients associated with the explanatory variables which are not identified (Holm & Jæger, 2011).
[36] As the dependent variable is binary, the error variance is unknown and a functional form for the underlying probability distribution is assumed. For probit models, the error variance is normalised to 1. For logit models, it is normalised to $\pi^2/3$ (Holm & Jæger, 2011).

The ability of probit selection models to overcome selection bias is highly dependent on data quality and the amount of information available in the data (Holm & Jæger, 2011). The data that have been used in some studies are particularly problematic in terms of accuracy of measurement and/or definition of variables of interest and comparability across data sets (see Jackson et al., 2007 for instance). The administrative data used here provide a census of Scottish state school pupil attainment in terms of formal qualifications over an eight-year period; therefore, there should be limited sample selection problems, certainly at the first transition. That said, clearly the data lack the richness and specificity of survey data in terms of detailed background information on individuals.

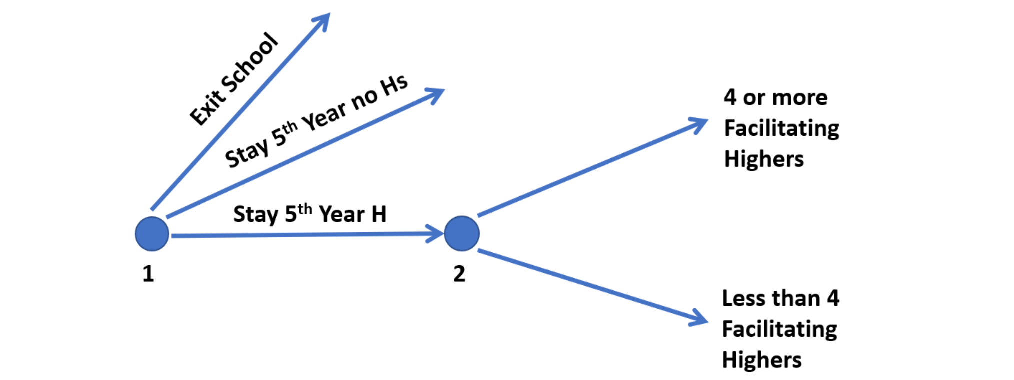
To summarize, the use of sequential logistic regressions with what are effectively census data, to model two educational transitions, provides a time-invariant approach to the analysis of secondary effects in educational choices that should not suffer from selection bias but may be subject to some OVB (as is the case with most models in the Education/Labour Economics' field). In this respect, it should be noted that modelling individuals' educational decisions in this manner makes the implicit assumption of unrestricted choice. In practice, individuals' educational decisions take place in a restricted choice environment. As noted in the Introduction, choices are restricted by the institutional framework in terms of the curriculum structure which is determined, in the first instance, by government policy to be delivered by local authorities through schools. Where there is choice within a set curriculum, this will be constrained at school level by staffing and timetabling restrictions. Over and above this, *individual* subject choices will be influenced by teacher guidance/decisions (perhaps driven by school league table concerns) and parental input.

1.4.2 Model Specifications

The age 16 educational choices modelled are depicted in Figure 1.2. The first choice point (1) is where individuals choose whether to stay on to Fifth Year (S5) to take at least one Higher (among a package of qualifications), or stay on to take other qualifications (but no Highers), or exit the secondary school system entirely to enter the labour market or go into FE. The second choice point (2) is where those that stay on to take Highers make their subject choices: taking four or more facilitating subjects to follow the top academic track route, or less than four facilitating subjects, thereby pursuing a lower academic route. The choices modelled are *on-time* choices, whereby individuals move linearly from one level of qualification to the next in their senior school years; that is all the Highers taken are sat in S5 (not S6).

Figure 1.2

Secondary School Age 16 Educational Decisions



Choices Modelled:
1 = Stay on for 5th Year takes 1 or more Highers / Stays but takes no Highers / Exits to FE or labour market
2 = Studies 4+ facilitating subjects at Higher / less than 4 facilitating subjects

Sequential logits are used to model these choices, specifically:

- the decision to *stay on* beyond compulsory school age and take at least one Higher – dependent variable: *Hstay* and
- the subsequent decision to take four or more Highers in facilitating subjects – dependent variable: *Hfs4*.

Both dependent variables had outcomes: 0 = no, 1 =-yes.

Hstay was broadly defined as staying on to take a minimum of one subject at Higher to ascertain the likelihood of individuals progressing to the next qualification level in their fifth year as opposed to mitigating low age-16 attainment. The reference category (outcome 0) was comprised of the individuals who either left the school system entirely or stayed on but did not sit any Highers in S5 (and, therefore, did not progress in terms of the level of qualification).

Hfs4 was defined more strictly, specifying four as the impactful level of uptake for Highers in facilitating subjects as this is the threshold level of attainment for entry to more prestigious courses (e.g Medicine, Law) and universities and has been found to be the average number of facilitating subjects taken by pupils from professional backgrounds, compared to two or three for those from working class backgrounds (Iannelli & Klein, 2016). The reference category (outcome 0) was comprised of those individuals who stayed on to S5 but studied for less than four facilitating subjects and, therefore, did not pursue the top academic track.

The sequential logit models estimated the log odds of staying on to take at least one Higher (Hstay) and, subsequently, taking four or more Highers in facilitating subjects (Hfs4) as specified below:

$$
\begin{aligned}
ln\left(\frac{p_i}{1-p_i}\right) = {} & \beta_0 + \beta_1 Sex_i + \beta_2 SGI\ UCAS\ points_i \\
& + \sum_{j=3}^{7} \beta_j SIMD1-5_i + \sum_{k=8}^{11} \beta_k School\ quartiles1-4_i \\
& + \sum_{l=12}^{16} \beta_l Sex_i * SIMD1-5_i + \sum_{m=17}^{20} \beta_m Sex_i * School\ quartiles1-4_i \\
& + \sum_{n=21}^{40} \beta_n School\ quartiles1-4_i * SIMD1-5_i + \beta_{41} SGI\ Cohort\ Size_i \\
& + \beta_{42}\%\ pupils\ SIMD1\&2_i + \beta_{43} Employment_i \\
& + \sum_{p=44}^{49} \beta_p UrbanRural_i + \sum_{q=50}^{81} \beta_q LA + \sum_{r=82}^{89} \beta_r Year + \varepsilon_i
\end{aligned}
$$

Where:

- Sex (Male=1)
- SGI UCAS points: total Standard Grade/Intermediate 2 UCAS points at age-16
- SIMD 1–5: individuals' household Scottish Index of Multiple Deprivation 2009 quintile
- School quartiles 1-4: individuals' SGI UCAS points' school quartile
- Sex*SIMD 1-5: interaction term for sex with individuals' SIMD quintile
- Sex*School quartiles 1-4: interaction term for sex with individuals' School quartile
- School quartiles 1-4*SIMD 1–5: interaction term for individuals' School quartile with SIMD
- SGI cohort size: school size as measured by Standard Grade/Intermediate 2 cohort size
- Higher cohort size: school size as measured by Higher cohort size

- % pupils SIMD 1 & 2: schools' socio-economic composition given by the percentage of pupils from SIMD quintiles 1 and 2, the most deprived 40% of households
- Employment: male/female youth employment rate in local authority area
- Urban/Rural: dummy variables for urban/rural location and degree of remoteness
- LA: local authority dummy variable for individuals' school
- Year: dummy variables for years 2002-2009

Summary statistics for the data used in both logit models are provided in Appendix A, Tables A1 and A2. The working sample for the first logit model (277,125), the decision to stay on to take at least one Higher, is lower than the total number of cases in the data (489,468) as a result of incomplete and/or inconsistent recording of candidate details when data are entered. For instance, failure to record candidate postcodes means that no SIMD background can be attributed, while failure to record a candidate's school means that school characteristics in terms of size and pupil composition cannot be attributed. The working sample for the second logit model (185,698), the decision to take four or more Highers in facilitating subjects, is further reduced by default, as this is confined to those that stayed on to take at least one Higher whilst also being subject to attrition resulting from data entry inconsistencies.

Whilst this is ostensibly population data, significance testing is reported for two main reasons. Firstly, the attrition caused by data entry inconsistencies makes this necessary. Secondly, whilst the data provide a census of attainment, it is a census of attainment for *those who were entered for assessment*. The data do not record those who may have started, for instance, an S5 programme of study and then dropped out before being entered for formal assessment. If this unknown attrition is to be considered as equivalent to failing formal assessments then, the number of failures is under-reported and arguably gives rise to the issue of measurement error and the potential for OVB.

School level cluster-robust standard errors are employed in both decision models to account for the possibility that model errors for individuals in the same school might be correlated (although model errors for individuals in different schools are not). Failing to control for within-cluster error correlation can produce misleadingly small standard errors, resulting in narrow confidence intervals, large t- statistics and low p- values and so lead to the conclusion that a variable is significant when, in fact, it is not (Cameron &

Miller, 2015). Standard errors that control for within-cluster correlation are found to be often several times larger than default standard errors where such correlation has been ignored (*ibid*).

1.4.2.1 Individual Characteristics

Clearly, prior attainment will influence individuals' decisions to both pursue education beyond compulsory schooling and the subsequent nature of that education. Individuals' prior attainment, and arguably academic ability, is measured by the derived variable SGI UCAS points that gives a total of an individual's UCAS points gained from their combined Standard Grade and Intermediate 2 performance at age 16; a UCAS tariff points' score. Standard Grades were being increasingly replaced by Intermediate 2 qualifications during the decade 2000-2009 and so it is necessary to consider attainment in both qualifications by age 16. SIMD 1-5 indicates the candidate's socio-economic background as measured by the Scottish Index of Multiple Deprivation 2009.

As well as an absolute measure of age 16 attainment, a relative measure is also included in the form of the Standard Grade and Intermediate 2 points' quartile within the school into which an individual falls - School quartiles 1-4. This is included to attempt to capture any varying attention and support that might be given by teaching staff within schools to candidates according to their abilities against the background of school league table pressures. The basic premise being that more able pupils may attract more teaching input/support to help them realise their potential in securing qualifications' grades/passes and thereby improving a school's league table standing. On the other hand, pupils of average ability may be encouraged to play safe in terms of subject choices and be directed towards subjects that are perceived as being 'easier' to pass. It is also possible that low achievers may receive extra support as schools try to increase their percentage of pupils reaching policy benchmark targets to increase league table position. The use of alternatively constructed points' quartiles based on local authority areas and Scotland-wide was explored. The impact was marginal on the overall fit of the sequential logit models (see 1.4.3 below, Predictive Power of Models) and it was decided to retain the school-based quartiles to try to capture differences in pupil treatment at this level. Interaction terms are included to explore possible relationships between sex and SIMD, sex and relative age-16 achievement (School quartiles 1-4) and, crucially, relative age-16 achievement and SIMD to ascertain whether or not a socio-economic effect is present within the same achievement levels.

1.4.2.2 School Characteristics

To model school-specific effects, the variables *SGI cohort size* (for staying on), *Higher cohort size* (for taking four or more Facilitating Subjects) and *% pupils in SIMDs 1 & 2* are included. The latter is the proportion of pupils from SIMD2009 quintiles 1 and 2 (the least well-off 40%) for each school in each year, giving an indication of the schools' socio-economic demographic. This is preferred to the inclusion of an SIMD2009 quintile dummy mapped to the schools' postcodes as pupils and schools' SIMD locations do not necessarily coincide. For instance, pupils resident in an SIMD2009 quintile 1 area (the most deprived 20%) may attend a school based in an SIMD2009 quintile 5 area (the least deprived 20%). The variables SGI cohort size and Higher cohort size measure the size of schools' Standard Grades/Intermediate 2 and Highers' cohorts in any given year, indicating not just school size but also indirectly, resources and subject availability. It might be expected, for example, that larger schools have more resources and are able to offer a greater choice of subjects for study. On the other hand, pupil/staff ratios might be lower in smaller schools but subject choice may be more limited.

1.4.2.3 Other Variables

In addition to school-specific effects, school policy will be dictated by local authority education goals and policies. To capture this, the dummy variable *LA* for school local authority is included. Also included are a set of dummy variables - *Urban/Rural* that indicate the nature of schools' geographical location as, for instance, individuals attending schools in large urban areas might be expected to have more educational choices than those at schools in remote rural areas.

Finally, *Year* - year dummies for the eight years 2002-2009 and *Employment* – the annual male/female youth employment rate (16-24 year olds) for the LA area (as appropriate) are included. The latter are ONS labour market data that were linked to the dataset by mapping to the LA areas. [37] It might be expected that youth employment in the local area

[37] Male and Female Youth employment statistics were missing for the years 2002 and 2003. These were imputed by weighting total male and female employment for 2002 and 2003 by the average percentage ratio of the male/female youth employment rate to the overall male/female employment rate for the

would influence the decision to stay on and take Highers; that is, the lower the level of youth employment, the more likely it is that individuals would choose to stay on to acquire further qualifications. The youth employment rate is used in preference to the general unemployment rate as the latter may be partially hidden by the tendency for school leavers who cannot find employment to enrol at FE colleges.

1.4.3 Predictive Power of Models

In terms of the choice of definition for an individual's SGI points quartile, that is whether this should be school-, LA- or nationally-based, there were negligible differences in the size of the log likelihoods for models containing these alternatives, suggesting negligible changes in their explanatory power over null models. For both logit decision models (the decision to stay on to take Highers and the decision to study four or more Facilitating Highers), the log likelihood rises as more aggregated points' quartiles are used (school to LA to Scotland-wide). For all the different quartile versions of the models, however, the pseudo-R2 remained at 54% for the first logit decision model and 38% for the second whilst the percentage of correctly classified observations was constant at 87% and 83% respectively. Given this, the school-based quartiles were used as discussed above (1.4.2.1 Individual Characteristics).

Testing the Hstay model of the decision to stay on at school and take Highers revealed that 87% of observations were correctly classified. Calculation on the Adjusted Count R^2 demonstrated a 69% reduction in the classification error rate compared to a prediction based only on the dependent variable's marginal distribution. The goodness-of-fit test based on the Pearson chi-squared statistic was significant at the 0.001 level.[38]

Local mean regressions were used to graphically analyse whether the relationship between the dependent variable and the continuous independent variables (SGI UCAS points, SGI cohort size/Higher cohort size, % pupils SIMD 1 & 2, Employment) was linear

later years. On average, the male youth employment rate was 80% of the total male employment rate for the years 2004-2009, while the female youth employment rate was 89% of the total female employment rate.

[38] The Pearson chi-squared statistic evaluates the sum of the square of the Pearson residuals across all covariate patterns; it tests the hypothesis of conformity between the predicted and observed frequencies across the covariate patterns (Kohler & Krueter, 2009). The Pearson residuals compare the number of successful classifications of individuals having particular covariate patterns with the predicted number of successes having those covariate patterns.

or not (Appendix A, Figures A1-A4). [39] Ideally, the local mean regressions would exhibit an S-shaped curve (probability distribution) or linearity on a small section of the S-shaped curve where range of the band means are small. For the Hstay model, the scatterplot of SGI UCAS points (Figure A1) suggests an S-shaped curve although there is some disturbance at the extreme end of the upper tail. The SGI cohort size and % pupils SIMD 1 & 2 variables both exhibit linearity (Figures A2 and A3). The size of the age 16 cohort (SGI cohort size) appears to be positively and linearly related to the likelihood of staying on to take Highers whilst a school's percentage of pupils from the most disadvantaged 40% of households (% pupils SIMD 1 & 2) exhibits a negative linear relationship to this. Male/female youth employment (Figure A4) exhibits a potentially problematic inverse u-shaped relationship but the extensive middle section of the curve appears to show a positive, shallow linear relationship between the level of youth employment and the probability of staying on to take Highers.

There are a number of outliers noticeably concentrated in the tails of the distribution of the residuals as shown by the scatterplots of: leverage against the standardised residuals (Appendix A, Figure A5) [40], the change in the coefficients against the predicted probabilities (Appendix A1, Figure A6), the change in the Pearson chi-squared statistic against the predicted probabilities (Appendix A, Figure A7) [41]. All of these suggest that the model may be less accurate in predicting the likelihood of staying on to take Highers at the extremes of the sample distribution; incorrectly predicting leaving at the lower end and staying on at the top end respectively.

For the Hfs4 model of the decision to take four or more Facilitating Subject Highers, 83% of observations were classified correctly. The Adjusted Count R^2 indicated a 40% reduction in the classification error rate compared to a prediction based only on the

[39] The median trace used as a scatterplot smoother for linear regression models cannot be used for dichotomous dependent variables as the median will have values of 0 or 1 only. Hence the use of local mean regressions instead.

[40] Leverage indicates observations that have an unusual combination of values for the independent variables. Standardised residuals are used to approximate discrepancy – an unusual characteristic of the dependent variable, given the values of independent variables.

[41] That is, the change in the Pearson Chi-squared statistic when the jth covariate pattern is removed from the dataset. Kohler & Kreuter (2009) suggest that the scatterplot of the change in the Pearson Chi-squared statistic against the predicted probabilities is well suited for the discovery of covariate patterns that are hard to predict through the model.

dependent variable's marginal distribution. The goodness-of-fit test (Pearson chi-squared statistic) was significant at the 0.001 level.

Graphical analysis of local mean regressions revealed an S-shaped relationship between studying four or more Facilitating Subject Highers and individuals' Standard Grade/Intermediate 2 points (SGI UCAS points) – (Appendix A, Figure A8). The S-curve probability distribution had a longer flatter tail at low values of SGI UCAS points than the staying on model but exhibited a similar sharp increase in outcome likelihood at the 200 points mark, indicating the importance of reaching a certain level of prior attainment. The size of an individual's school Highers' cohort (Higher cohort size) exhibited a positive linear relationship with the probability of choosing four plus Facilitating Highers whilst the percentage of pupils from the most disadvantaged 40% of households (% pupils SIMD 1 & 2) in a school suggested a negative linear relationship (Appendix A, Figures A9 and A10 respectively). The plot for male/female youth employment was flat, suggesting no relationship between local employment levels for young people and their choice of Higher subjects (Appendix A, Figure A11).

For the Hfs4 model, outliers were concentrated in the lower tail of the distribution of the residuals (see scatterplots of: leverage against the standardised residuals - Appendix A, Figure A12, the change in the coefficients against the predicted probabilities - Appendix A, Figure A13, the change in the Pearson chi-squared statistic against the predicted probabilities - Appendix A, Figure A14). This would suggest that the model may be less accurate in predicting the likelihood of studying four or more Facilitating Highers at lower end of the sample distribution; incorrectly predicting the uptake of Facilitating Highers.

The Likelihood Ratio Tests and Post Estimation Tests of coefficients (where appropriate) indicated that all the independent variables were significant at the 1% level for both logit models with one exception; male/female youth employment in the Hfs4 model (decision to take four or more Higher Facilitating Subjects) was insignificant (Appendix A, Table A3). For the Hstay model (decision to stay on to take Highers), exclusion of the year dummies and youth employment produced the greatest falls in the Pseudo R^2 from 54% to 41%/40% respectively but no similar impact was seen in the Hfs4 model (where, as above, youth employment was insignificant). This would suggest that wider economic and environmental factors play a part in the initial decision to stay on to take Highers but not in the subsequent subject choice.

In both decision models, the exclusion of the SGI UCAS points variable (individuals' age-16 UCAS tariff points' score) results in noticeably lower maximised values of the likelihood functions and the pseudo R^2 falls by 7% for staying on to take Highers and 20% for taking 4 plus Higher Facilitating Subjects (Appendix A, Table A3). This indicates that, as might be expected, age-16 attainment is the main determinant of staying on to take Highers and, particularly, taking the necessary number of Facilitating Highers for entry to more prestigious universities.

1.5 Results and Analysis

In general, as a minimum of four or five facilitating subjects was effectively compulsory pre-CfE, the majority of fourth year pupils would have been studying for SG/Intermediate 2 qualifications (at the various levels) in English, Maths, a science, a modern language and a humanity (e.g. History or Geography). Pupils had to make closed choices in these latter three areas; for example, choosing whether to take Biology or Chemistry or Physics or general Science. In addition to this, state secondary schools usually offered provision in a *technical* subject (e.g. Computing or Graphics), a *creative* subject (e.g. Art or Physical Education) and finally an open choice whereby individuals could choose to take a second science, modern language, humanity or another subject.

Between 2002 and 2009, 60% of pupils in state secondary schools studied for 5 or more SG/Intermediate 2s in facilitating subjects (Figure 1.3). When this is examined by SIMD 2009 quintiles (Figure 1.4), there is no significant difference in the uptake of Facilitating SGIs at the 5 subjects' mark. Once this is passed and pupils were able to exercise choice, it can be seen that those from the most deprived 40% of households were significantly less likely to be studying more than five facilitating subject SGIs compared to those from the other, less deprived quintiles.

In general, where there is choice over subject uptake for formal examinations at age 16, those from more disadvantaged backgrounds are significantly less likely to choose Facilitating Subjects. For example, of those taking a second science, just over 10% were from the most disadvantaged twenty percent of the state school population compared to almost 30% from the most advantaged twenty percent (Figure 1.5). Distributions of the uptake for second language and humanities' choices, as well as English and Maths, can be seen in Apprendix A, Figures A.15 to A.17.

Figure 1.3

Percentage of Pupils Studying Facilitating Subjects in S4

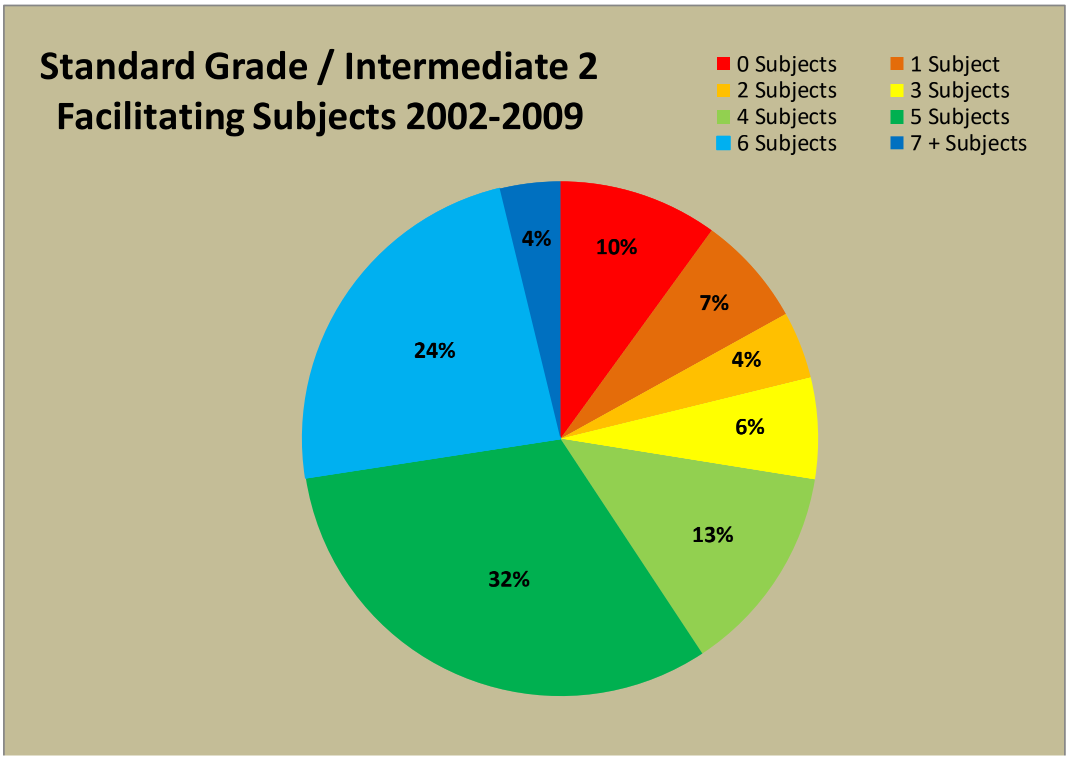
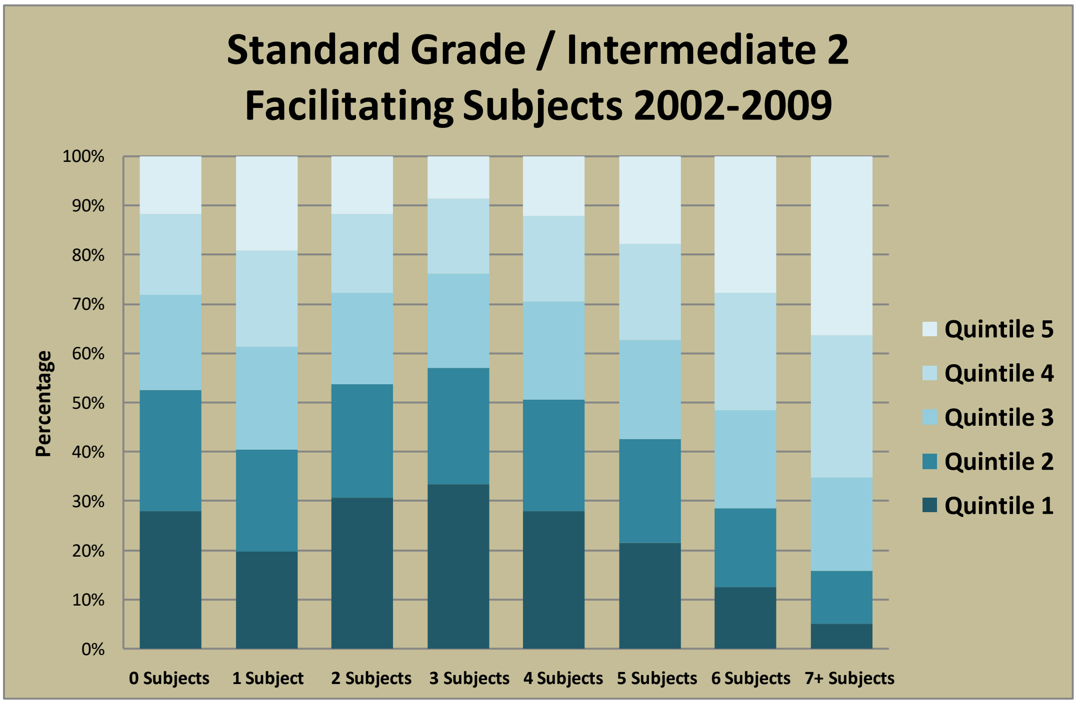


Figure 1.4

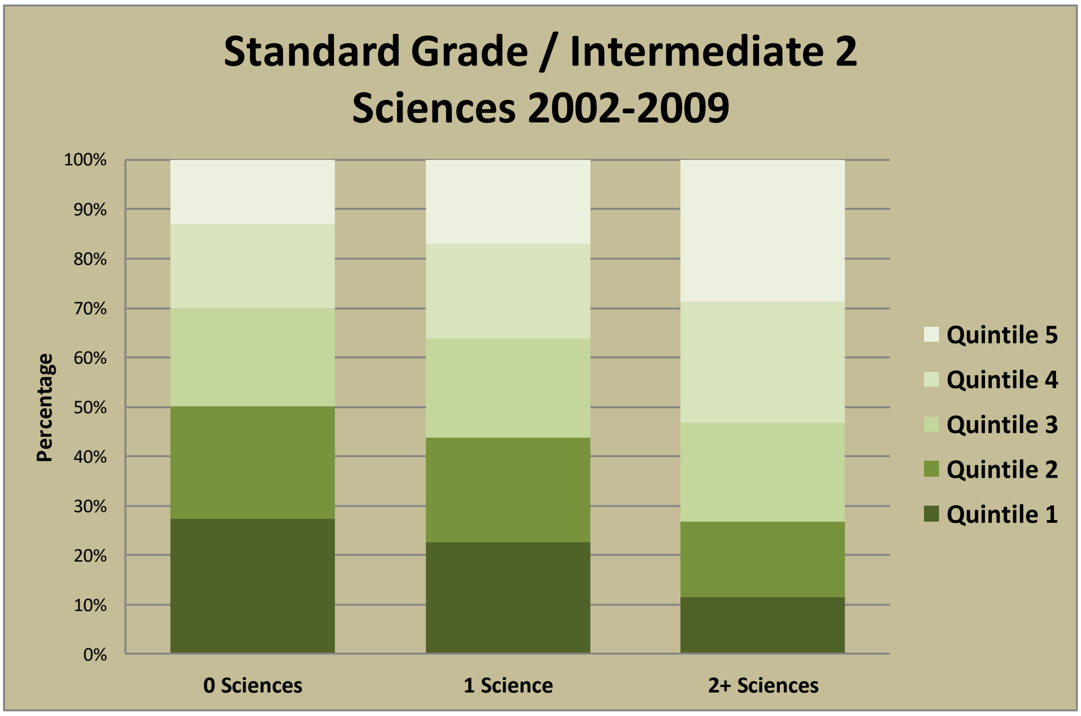
Number of Facilitating Subjects Studied in S4 by SIMD



Pearson χ^2 (28) = 30000   Pr = 0.000

Figure 1.5

Number of Science Subjects Studied in S4 by SIMD



Pearson χ^2 (12) = 23000   Pr = 0.000

1.5.1 Staying on to take Highers

In the logit analysis of the decision to either stay on to take at least one Higher, sex, SIMD quintile and age-16 achievement (SGI points) were all found to be significant both in their own right and in most of the interactions between them (Table 1.1). [42] Males were less likely to stay on than females to take Highers; a male student being only 79% as likely to stay on as a female student. Pupils are increasingly less likely to stay on, the lower their SIMD quintile when compared with the top quintile 5 (the least deprived or most affluent). Those from the most deprived 20% of households were only 54% as likely to stay on as those from the most affluent 20%. For one extra Stand Grade/Intermediate 2 UCAS point gained, the likelihood of staying on to take Highers increased, on average, by 3%.

Individuals' school-based, SGI UCAS points' quartiles (School quartiles 1-4) were all significant. Those in quartiles 2 and 3 were less likely (61% and 81% respectively) to stay on compared to those in the highest points' (top) quartile 4. Individuals in the lowest points quartile 1, however, appeared to be almost 10 times as likely to stay on to take Highers as an individual in the top quartile. This effect, however, is largely mitigated by coming from SIMD 1 (the most deprived 20% of households) as shown by the interaction of relative attainment (School quartiles 1-4) with SIMD. [43] This result is in keeping with Croxford's SSLS findings for 1985-2005 (Croxford, 2009) of a marked increase in staying on at school after age 16 among low attaining groups, perhaps because of the influence of the Higher Still reforms combined with the EMA. The EMA was piloted in 1999 and rolled out nationally in the academic year 2004/5. This may be reflected in the broad definition of the dependent variable (Hstay) which is defined on the basis of staying on to take at least one Higher in any subject. It is possible that those with a very poor age 16 performance do not have any real choice other than to stay on to take further qualifications and contribute to the dependent variable by taking a single Higher and increasing the numbers recording a '1' for this. If so, this single Higher is likely to be part of a package with other lower secondary qualifications to try to repair poor S4 attainment.

[42] Table 1.1 provides key results. The full set of results for this model can be found in Appendix A, Table A4.

[43] The respective logit coefficient values were 2.28 and -1.92.

Table 1.1

Sequential Logit Analysis:

	Decision to Stay on to take at least One Higher	Decision to take Four or more Facilitating Highers
Sex (M=1)	0.791***	1.053
SGI UCAS points	1.034***	1.045***
SIMD 1	0.540***	0.740***
SIMD 2	0.638***	0.822***
SIMD 3	0.686***	0.860***
SIMD 4	0.868*	0.923**
School quartile 1	9.796***	3.907
School quartile 2	0.610***	0.899
School quartile 3	0.806**	0.684***
Sex*SIMD 1	1.082	0.937
SIMD 2	1.008	0.922
SIMD 3	1.000	1.021
SIMD 4	0.912	1.002
Sex*School quartile 1	0.658***	3.394
School quartile 2	0.800***	0.928
School quartile 3	0.971	1.063
School quartile 1*SIMD 1	0.147***	14.288*
SIMD 2	0.184***	19.472
SIMD 3	0.372***	0.316
SIMD 4	0.443***	1.084
School quartile 2*SIMD 1	0.518***	3.120
SIMD 2	0.502***	3.086**
SIMD 3	0.682***	0.910
SIMD 4	0.719***	0.771
School quartile 3*SIMD 1	0.757***	0.803
SIMD 2	0.818**	0.812*
SIMD 3	0.895	0.781**
SIMD 4	0.861**	0.893
SGI cohort size	0.999	-
Higher cohort size	-	1.001
% pupils SIMD 1 & 2	0.994***	0.994*
Employment	0.992***	-
Observations	277125	185698
Log pseudolikelihood	-86194.425	-67545.4
Wald chi2(73)	25021.71***	15662.55***
Pseudo R2	0.541	0.388

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles
Omitted reference categories: SIMD 5 – the least deprived (most affluent) 20% of households, School quartile 4 – top school SGI UCAS points' quartile
Other independent variables included: Urban/Rural location, Local Authority and Year fixed effects
Cluster-robust standard errors at the school level (Model 1: 352, Model 2: 348)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The top age-16 achievers, on the other hand, may have more choices open to them; for instance, options in terms of employment or FE courses.

The interaction terms for sex and the SGI UCAS points' quartiles indicated that lower attaining males (in the bottom two quartiles) were significantly less likely to stay on to take Highers than those in the top quartile. The points' quartiles/SIMD quintiles' interaction terms reveal the influence of socio-economic background *within* the pattern of age 16 achievement. All points' quartiles/SIMD quintiles' interaction terms were significant. In general, it can be seen that the likelihood of an individual staying on to take at least one Higher rises with SIMD quintile at the different levels of achievement. For instance, an individual located in the lowest points' quartile (School quartile 1) and SIMD quintile 1 (the least well-off 20% of households), is only 15% as likely to stay on to take Highers compared to an individual from the top quartile 4/highest SIMD quintile 5; whereas an individual in the lowest points' quartile 1 and SIMD quintile 4 is 44% as likely.

Schools' socio-economic composition was significant while school size (SGI cohort size) was not. For a one unit increase in a school's the proportion of pupils from the least well-off 40% of households (% pupils SIMD 1 & 2), the odds of staying on to take Highers fell on average by 1%. There were no significant differences across urban or rural schools (see Appendix A, Table A4) and no obvious geographic pattern was apparent in terms of LA areas (Appendix A, Table A5 shows the LA dummy variables sorted by coefficient size/odds ratios). Glasgow City, Scotland's largest local authority and, therefore, the area in which there is the greatest school/FE and subject choice, was omitted as the reference category. There were only two LA areas, East Renfrewshire, and Eilean Siar where pupils were more likely to stay on to take Highers than those in Glasgow schools. Those in East Renfrewshire and Eilean Siar LA schools were twice as likely as Glasgow school pupils to stay on to take Highers. Pupils in schools in six LAs were less likely to stay on to take Highers than individuals in Glasgow schools; with those in Moray, Aberdeenshire and the Shetland Islands being less than 50% as likely to do so. This may be a reflection of the very different nature of these rural economies whereby, having completed compulsory education, the traditional expectation is that a young person leaves school to obtain work locally. The remaining 23 LAs exhibited no significant differences from Glasgow.

Year dummies for 2002 to 2006 (Appendix A, Table A4) were all significant suggesting that individuals were between 11% and 17% less likely to stay on to take Highers in these years; 2007 and 2008 were not significantly different from 2009 (the reference year). The odds fell from 2002 and then start to rise again in 2006; this may reflect the influence of the Higher Still policy initiative and a generally increasing emphasis on credentialism. It might be expected that rising youth employment would reduce the likelihood of an individual staying on to take Highers. The impact of male/female youth employment is significant, indicating that for every 1% rise, the likelihood of an individual staying on to take Highers was reduced on average by 1%.

1.5.2 Taking Four or more Facilitating Subject Highers

The sequential logit analysis of the uptake of four or more Highers in Facilitating Subjects (**Table 1.1**) again showed SIMD quintile and age 16 achievement to be significant determinants but the main sex effect was insignificant as were all the related interactions terms (sex/SIMD and sex/SGI points' quartiles). [44] Significant individual SIMD quintile effects show that those from more deprived households were less likely to study for four plus Facilitating Highers. For instance, an individual from SIMD 1 (the most deprived 20% of households) was 26% less likely to study four plus facilitating Highers compared to someone from SIMD 5 (the least deprived 20% of households). The odds of taking four or more Facilitating Highers increase, on average, by 5% for an extra Standard Grade/Intermediate 2 UCAS point gained. Of the school-based, SGI points' quartiles, quartile 3 was significant suggesting that those in the upper quartile were only 68% as likely to take four plus Facilitating Highers as those in the top quartile. Quartile 1 was significant also (at the 10% level) implying that those in this lowest quartile were four times as likely to take four or more facilitating Highers compared to someone in the top SGI points' quartile.

The interaction terms between the upper points' quartile and SIMD provide some evidence of a secondary inequality effect as students from SIMD quintiles 2 and 3 households with 'good' above median attainment at age 16 were found to be approximately 20% less likely to take four plus facilitating Highers than compared to those

[44] Table 1.1 provides key results. The full set of results for this model can be found in Appendix A, Table A4.

with equivalent attainment from more affluent backgrounds – quintiles 4 and 5. For the bottom points' quartile, the SIMD interaction terms indicated that those from the most deprived 20% of households were more likely to take four plus facilitating Highers than individuals in the top points' quartile from the least deprived 20% of households. This result probably reflects the impact of outliers as only 5% of individuals taking four or more facilitating Highers came from SIMD 1 households.

Schools' socio-economic composition (% pupils SIMD 1 & 2 – the proportion of pupils from the least well-off 40% of households) was significant once again while school size was not. For a one percent increase in the proportion of pupils from the least well off 40% of households in a school, the odds of an individual staying on to take four plus Facilitating Highers fell on average by 1%. The urban/rural location of schools was again insignificant. Only pupils in Eilean Siar schools were more likely to take four plus Facilitating Highers than those in Glasgow schools, where they were 63% more likely to do so (Appendix A, Table A4). In schools in 20 other LAs, pupils were less likely to take four or more facilitating Highers compared to those in Glasgow schools; with pupils in Aberdeenshire and Moray being only 40% and 27% respectively as likely to do so. Again, the latter may be a reflection of both subject availability and very different local economies. Schools in the remaining 10 LAs were insignificantly different from those in Glasgow. There was no particularly strong correlation between the LAs' ranked differences (compared with Glasgow) in pupils' propensities to stay on to take Highers and then to take four or more facilitating Highers (Appendix A, Table A6). That is, the likelihood of choosing to study four plus facilitating Highers in an LA area is not strongly associated with the likelihood of staying on to take Highers; although there appears to be some small consistency at the top and bottom with an increased likelihood of both for pupils in Eilean Siar schools and a reduced likelihood of both for those in Aberdeenshire and Moray schools.

Year dummies from 2002 to 2004 were all significant indicating that individuals were significantly more likely to study four plus Facilitating Higher subjects the further back in time they were making their decisions (Appendix A, Table A4). Whilst the years 2005, 2006 and 2007 were not significant, 2008 was. The decreasing likelihood of studying four or more Facilitating Subjects from 2002-2004 may be explained by the increasing range of subjects available for Higher study throughout the decade. While what appears to be a break or potential reversal of the pattern in 2008, might be explained by an increased awareness of the benefits of studying Facilitating Subjects through, for instance, the provision of admissions' advice by the Russell Group.

To put the results in Table 1.1 into an overall context, consider what these might mean for two pairs of individuals with a male and female in each pair. Assume that one pair faces the most adverse set of characteristics/circumstances whilst the second pair faces the next most advantageous set of characteristics/circumstances compared to individuals in the reference category (who have the most favourable characteristics/circumstances). That is, pair one is from SIMD 1 and their SGI performance (assumed to be Foundation SG with 88 points) places them in the lowest points' quartile in their school. They are also assumed to be in the worst location and LA area (an accessible small town in Moray in this case) in terms of impact on the probability of taking four or more Highers in facilitating subjects. Pair two is from SIMD 4 and their SGI performance (assumed to be a mix of Credit SG 1 and 2s with 264 points) places them in the upper points' quartile in their school. They are assumed to be in the best location and local authority area (an accessible rural area in Eilean Siar) in terms of impact on the probability of taking four or more Highers in facilitating subjects. Summing the coefficients for their sets of independent variables and converting these into odds ratios indicates the overall degree of disadvantage and, within this, any gender effect. [45] For pair one, the likelihood of taking four or more Highers in facilitating subjects is just 9% for the male and 3% for the female compared to individuals based in Glasgow, from an SIMD 5 background and in their school's top points' quartile. Pair two, on the other hand, were 29% (male) and 26% (female) as likely. The gap between both pairs and the reference group, caused by relative deprivation, is stark. The difference between the two pairs is clearly substantial also and suggests the existence of a gender gap that closes as disadvantage lessens. For the most disadvantaged pair, although the likelihood of studying 4 plus facilitating Highers for the male is very low, it is three times that for the female, whereas the male/female ratio for the more advantaged pair is 1.1.

[45] This assumes that all the independent variables are jointly significant.

It is clear from the sequential logit analysis that both the decision to stay on at school to take Highers and the subsequent decision to study four or more facilitating subject Highers are significantly correlated with pupils' age-16 academic performance (primary effects) but that within this, socio-economic factors are influential too. Examining the linearity of age-16 performance with respect to taking four plus facilitating Highers (Appendix A, Figure A8), it can be seen that there is a long tail where SGI points are low. In both decision models, the exclusion of the SGI points variable results in significantly lower maximised values of the likelihood functions and the pseudo-R^2 falls by 7% for staying on to take Highers and by 20% (halving) for taking 4 plus Higher Facilitating Subjects (Appendix A, Table A3). This suggests that, as might be expected, age-16 attainment is the main determinant of staying on to take Highers and, particularly, taking the necessary number of Facilitating Highers for entry to more prestigious universities.

1.6 Conclusions

The educational transitions' literature makes the distinction between primary inequalities – differential attainment that occurs because of differences in family economic and social capital resources, and secondary inequalities – differences in the educational choices made by individuals from different socio-economic groups even when they have the same level of attainment. Much existing research and policy formulation has tended to focus on socio-economic differences in educational attainment, 'primary effects', in terms of the number of awards achieved at a given level; in particular, at age-16, the end of compulsory education in the UK. Individuals' life chances in terms of accessing tertiary education and their labour market outcomes, however, are *subject* choice dependent also. Some subjects carry more weight than others for both HE and FE entry and with employers. Achieving an 'A' grade Higher in Physical Education, for instance, is not the same as securing an 'A' grade Higher in Mathematics for either employment opportunities or university entry. The prestigious Russell Group of universities, often criticised for their low intake of individuals from more disadvantaged backgrounds, now issues guidance as to which subjects prospective applicants should study in school to *facilitate* access to their institutions. Essentially, Russell Group facilitating subjects are traditional academic subjects: English, Maths, Geography, History, Modern Languages and the named Sciences.

Research across the UK's education systems indicates that educational inequalities over the last 50 years have declined in terms of broad attainment at age 16 (with respect to numbers of qualifications achieved) and continued participation in education beyond the end of compulsory schooling. At the age-18 transition point to HE, however, inequalities in the allocation of university education would appeared to have strengthened with those from more disadvantaged households being less likely to gain entry to the more prestigious institutions and/or courses. In the UK and further afield, research has shown family characteristics to dominate financial constraints and expectations of successful completion of courses to be important in making educational decisions. In Germany, the perceived probability of success has been shown to be determined by children's grade points and type of school recommended at the end of the primary stage.

In Scotland, recent work using survey data has shown secondary inequalities in subject choices to explain differential HE entry. Given that children from disadvantaged backgrounds tend to perform less well at school, it is possible that child and parental educational expectations, formed on the basis of this lower achievement, may feed through into poorer, less challenging, secondary school subject choices thereby producing and sustaining inequalities in HE entry over time. It is likely also that children from more disadvantaged households will have less social capital to draw on in terms of immediate family members with knowledge of how the later levels of the education system operate and, if so, this might compound poor subject choice decisions.

This chapter examined facilitating subject choice in state secondary schools in Scotland from 2002 to 2009 using SQA administrative data that effectively provide a national census of qualifications' achievement. The main focus was the educational transition to the study of academic subjects at upper secondary level, the critical level of qualification for HE entry. The Mare sequential logit approach was adopted to model, firstly, the decision to stay on at school after age 16 and take at least one Higher and, secondly, the decision to study four or more Highers in facilitating subjects; that is the minimum number of academic subjects required for entry to more prestigious HE institutions and/or courses. This approach was chosen because it is time invariant to the increased distribution of upper secondary education in these years, resulting from the Higher Still reforms, and so enables the analysis of the underlying pattern of the allocation of secondary education at the post-compulsory level. Any selection bias is likely to be limited as only two decisions, made in quick succession, are examined. To the extent that this may exist, the comparative methods' transitions literature suggests that any bias is likely to be downwards, understating the effects of socio-economic background.

In general, the pre-CfE structure of the Scottish Education System mitigated against inequality in the choice of Facilitating Subjects for age-16 qualifications; five out of eight subjects studied were required to be traditional academic ones. The tendency for such inequality to appear was apparent, however, when pupils were faced with open choices. Individuals from more deprived households were seen to be less likely to choose to study, for instance, a second science or modern language. Secondary effects were evident across the different socio-economic groups, as defined by SIMD quintile, in both the decision to continue to stay on beyond the end of compulsory education and the choice of facilitating subjects at Higher. Individuals from SIMD 1, the most deprived 20% of households, were found to be only 54% as likely to stay on at school to take at least one Higher compared to those from SIMD 5 (the most affluent 20% of households). The analysis of the subject choice decision to take four or more Facilitating Highers revealed that individuals from the least well-off 20% of households were 26% less likely to study four or more Facilitating Highers compared to those from the most affluent 20% of households; this was despite their having similar academic achievement at age 16.

This analysis of SQA administrative data has added to the body of empirical work in the field by confirming the existence of secondary effects in subject choice at the upper secondary level in state schools Scotland-wide. It has demonstrated the potential for disadvantage in subject choice in Scottish secondary schools between the ages of 16 and 18; the critical period for the attainment of formal qualifications to progress to HE or FE or, indeed, enter the job market straight from school. Inequality has been shown to exist at the 'intensive' educational margin (rather than the 'extensive') with differential allocation across different socio-economic groups despite the general increase in distribution and, therein, attainment. The implication for policy is clear; closing the age-16 raw attainment gap in terms of a bald number of awards is not enough: subject choices need to be monitored. Equivalent attainment at age 16 does not appear to lead to equivalent aspirations in terms of post-compulsory school educational choices as measured by the uptake of Facilitating Subjects at Higher level. These results suggest that there is a need for improved information and guidance for both pupils and parents with respect to senior school subject choices. The need to monitor subject choice is arguably more acute now in the context of CfE and its ‘looser’ structure of subject and/or level of assessment choices (for when pupils are ‘ready’). This implies the potential for a multiplicity of educational tracks and, with this, a greater risk of maintaining and entrenching inequality through secondary effects as a result of imperfect information and uninformed decision making.

Chapter Two

Attainment in Facilitating Subjects

2.1 Abstract

Attainment in Scottish Secondary education is characterised by entrenched socio-economic and gender gaps. Pupils from the most deprived 20% of households are significantly less likely to achieve the benchmark attainment thresholds set by policymakers in terms of the number of awards gained at given qualification levels. In general, females have been seen to outperform males in secondary education since the 1970s. For progression from secondary to tertiary education, the subjects studied for formal qualifications are important too, not just the number of awards and/or grades. Some subjects carry more weight than others as they facilitate entry to more prestigious universities and degree programmes that attract higher labour market premia and social status. This chapter investigated the influence of gender and socio-economic background on attainment specifically in these so-called *facilitating subjects*: English, Maths, Geography, History, Modern Studies, Modern Languages, Biology, Chemistry, Physics. Multinomial logit models were estimated for each subject, at each qualification level, using SQA data from 2002 to 2009, to examine *within* subject attainment in terms of the likelihood of achieving either a low, middle or high pass as opposed to a fail. The impact of socio-economic background was found to be greater than that of gender. Children's relative ability was clearly important in terms of securing low passes but not strong enough to overcome socio-economic disadvantage to achieve higher level grades. The effects were particularly stark at Standard Grade/Intermediate 2, with the likelihood of securing any pass grade, in any subject, falling dramatically as socio-economic disadvantage increased sequentially from the least deprived 20% of households to the most deprived 20%. At Higher, socio-economic effects were found to be much reduced at low pass in all subjects but increased as pass grades rose. In general, males were outperformed by females in most subjects across the different qualification levels. Notable exceptions to this, however, were Maths and named sciences at Higher, the critical level of qualification for university entry, where males were seen to be significantly more likely to pass these subjects at all pass grades.

2.2 Introduction and Policy Context

It is clear from Chapter One that individuals from the most disadvantaged 40% of households are less likely to study Facilitating Subjects at school and, as a consequence, are likely to have narrower post-school education options and potentially less lucrative employment opportunities open to them. Whilst subject choice limits post-school education and employment choices, recent policy is still very much focussed on reducing the poverty-related educational attainment gap in terms of numbers of qualifications achieved, specifically between individuals from the least disadvantaged (most affluent) 20% of households and those from the most disadvantaged (most deprived) 20% of households.

There is much evidence to show that the poverty-related attainment gap in Scotland increases as children progress through the educational system. Children from deprived backgrounds leave school with significantly lowers levels of attainment than those from more affluent households, impacting directly on their post-school destinations and potential future earnings (Sosu & Ellis, 2014). Some evidence indicates that the attainment gap exists by age three and widens by age five in the areas of vocabulary and problem-solving whereby children from high-income backgrounds significantly outperform those from deprived households. By age five, differences in vocabulary and problem-solving scores between children from the least deprived and most deprived backgrounds were estimated to be equivalent to a 13-month and 10-month gap in development respectively (Bradshaw, 2011 in Sosu & Ellis, 2014). The latest results from the (now discontinued) annual Scottish Survey of Literacy and Numeracy (SSLN), for primary school levels P4, P7 and secondary school level S2 (ages 8-9, 11-12 and 13-14 respectively), show that attainment is clearly stratified by deprivation level at these key stages. The pattern of performance in literacy (reading, writing, listening and talking attainment) was found to be stable across the three surveys for 2012, 2014 and 2016, with children from the least deprived backgrounds outperforming those from the most deprived with, in general, no change in gap sizes (Scottish Government, 2017d). Across the three numeracy surveys (2011, 2013, 2015), the performance gap between least and most deprived pupils increased at P4 but remained the same for P7 and S2 pupils (Scottish Government, 2016). Expected gender patterns of attainment were also evident with boys tending to outperform girls in numeracy at S2 and girls outperforming boys in reading and writing. At S4, the end of compulsory schooling, from 2007 to 2012, there was a consistent gap in the average tariff scores of children from the least and most deprived backgrounds of some 300 points. [46]

The Scottish Government has committed itself to, not just closing this deprivation-related attainment gap, but substantially eliminating it during the next decade (Scottish Government, 2017a). Its National Improvement Framework (NIF) and Improvement Plan (*ibid*) aim to improve both *excellence and equity* by simultaneously raising attainment in general and specifically closing the attainment gap. The NIF focusses on literacy, numeracy, health and wellbeing across the 3-18 age range, listing the following key priorities:

- Improvement in attainment, particularly in literacy and numeracy
- Closing the attainment gap between the most and least disadvantaged children and young people
- Improvement in children and young people's health and wellbeing
- Improvement in employability skills and sustained, positive school-leaver destinations for all young people

Table 2.1 shows the 11 key measures by which progress is to be assessed and the current percentage point gaps between the proportion of individuals from SIMD 1 (the most deprived 20% of households) and those from SIMD 5 (the most affluent 20% of households) in the various education and health categories. Attainment in secondary school is to be measured by the qualifications young people have gained by the time they leave school. The attainment gap is defined as the difference in the percentage of individuals from SIMD quintiles 1 and 5 leaving school with variously:

- one or more SCQF level 4 qualification or above
- one or more SCQF level 5 qualification or above
- one or more SCQF level 6 qualification or above

[46] The tariff score is calculated by adding together all the grades (converted into tariff points) accumulated from all the different course levels and awards achieved by a student.

The data show that the attainment gap between most and least disadvantaged increases as qualification levels rise. SCQF levels are used to provide a broader measure than one based on specific SQA qualifications alone, so that, for instance, foundation apprenticeships can be included (Scottish Government, 2017a).

Table 2.1

Attainment Gap Key Measures

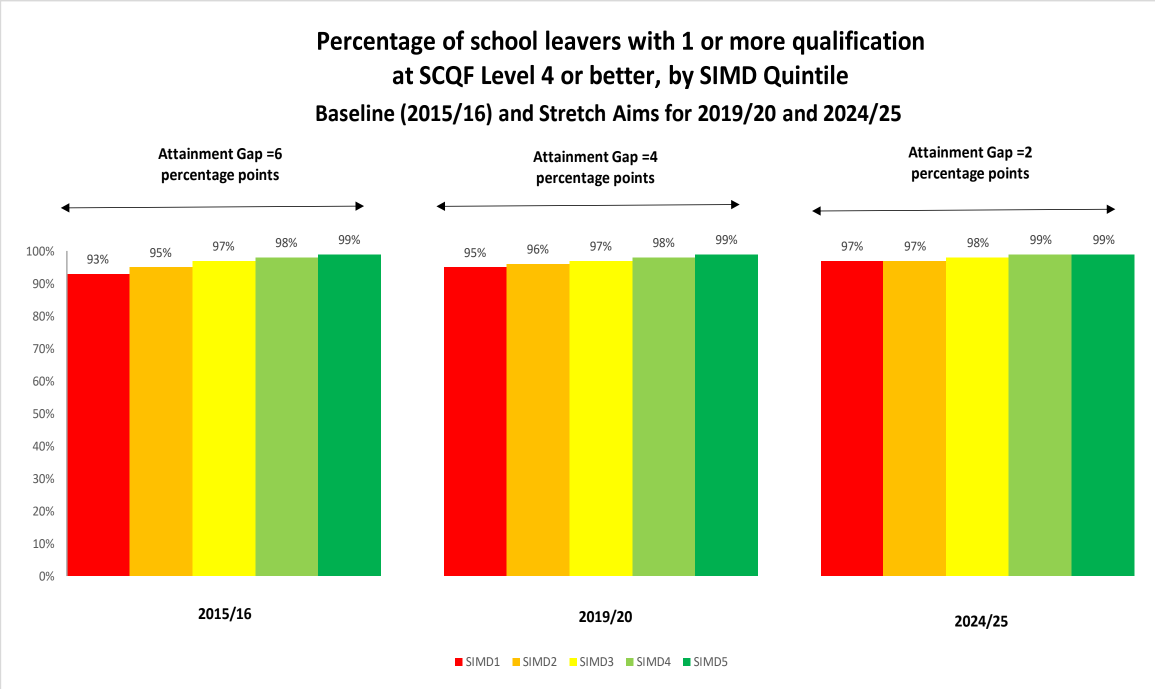
Measure	All Children %	Most disadvantaged (SIMD 1) %	Least disadvantaged (SIMD 5) %	Gap (percentage points)
27-30 month review (Children showing no concerns across all domains)	63.7	54.8	71.7	16.8
HWB: Children total difficulties score (age 4-12)	14	22	6	16
HWB: Children total difficulties score (age 13 & 15)	31	34	26	8
Primary – Literacy (P1, P4, P7 combined)	69.2	59.8	81.5	21.8
Secondary – Literacy (S3, 3rd level or better)	87.1	80.8	94.4	13.6
Primary – Numeracy (P1, P4, P7 combined)	76.4	69.2	86.5	17.3
Secondary – Numeracy (S3, 3rd level or better)	88.2	80.7	95.5	14.8
SCQF 4 or above (1 or more on leaving school)	96.3	92.8	98.8	6
SCQF 5 or above (1 or more on leaving school)	85.6	74.4	94.7	20.3
SCQF 6 or above (1 or more on leaving school)	61.7	42.7	81.2	38.5
Participation measure	91.1	84.8	96.3	11.5

(Source: Scottish Government, 2017a)

Figure 2.1 shows the 2016 position for all SIMD quintiles in terms of SCQF level attainment upon leaving school as well as the relative participation in post-school education, employment or training for SIMDs 1 and 5 and the so-called *Stretch Aims* of closing these gaps whilst raising attainment overall. By the middle of the next decade, the specific SIMD 1/SIMD 5 gaps are to be closed from: 6% to 2% at SCQF level 4, 20% to 5% at SCQF level 5 and 38.5% to 15% at SCQF level 6.

Figure 2.1

Attainment Gap SCQF Measures



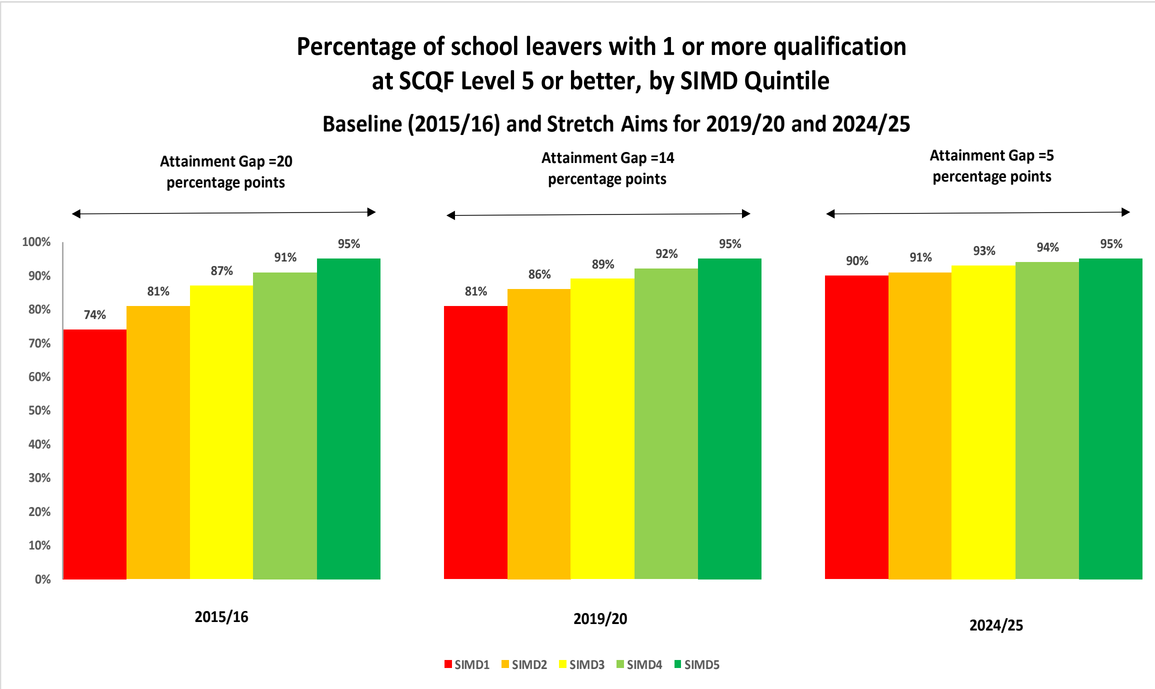
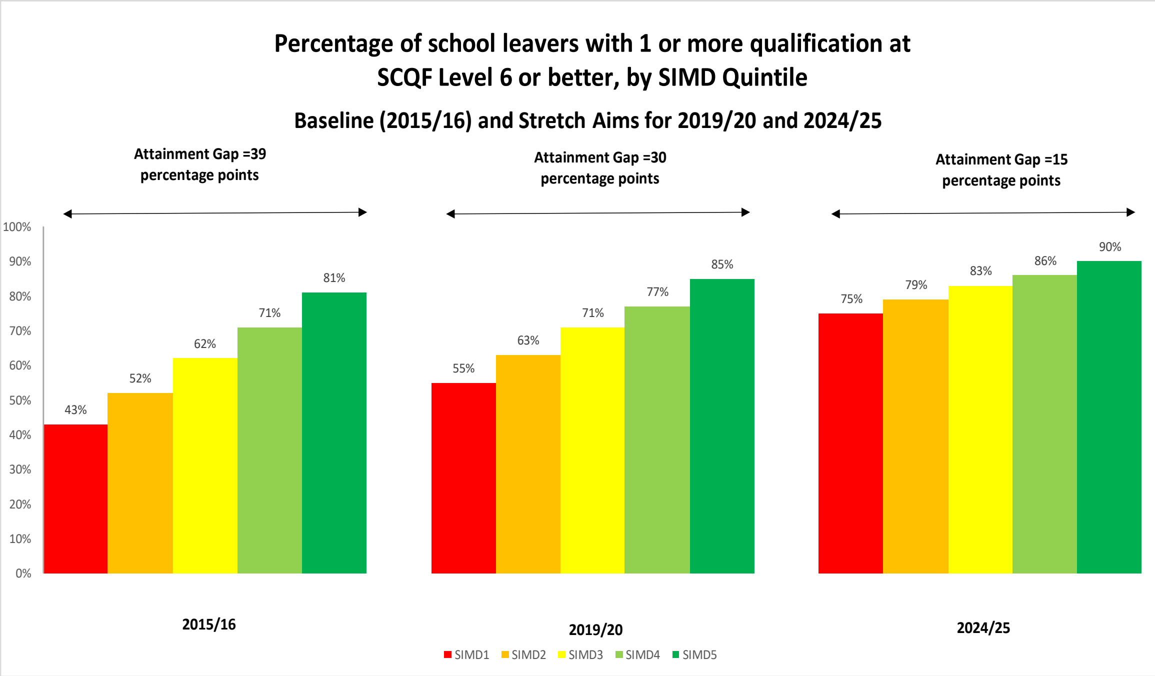
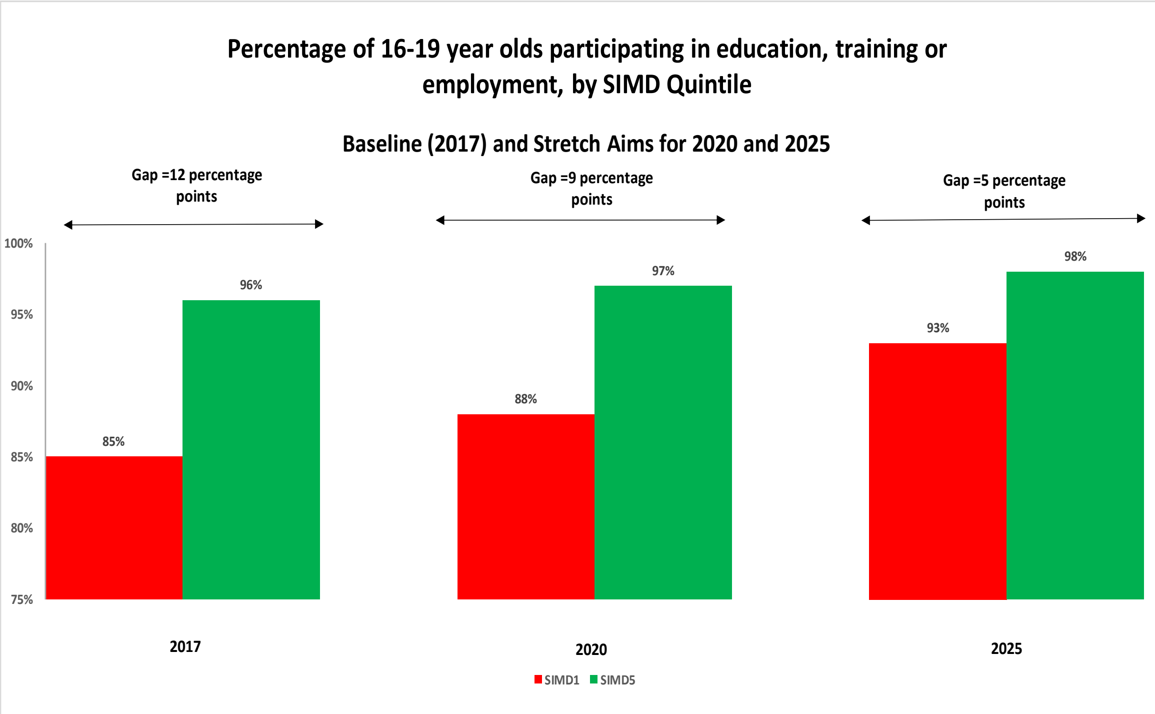


Figure 2.1 cont.

Attainment Gap SCQF Measures





(Source: Scottish Government, 2017a)

Scottish Government concern about geographical differences and high concentrations of deprivation in particular areas lead to the launch of the Scottish Attainment Challenge (SAC) at the beginning of 2015. Underpinned by the NIF, CfE and Getting it Right for Every Child (GIRFEC), the SAC aims to achieve equity in educational outcomes, by supporting 'schools and local authorities to focus on and accelerate targeted improvement activity in literacy, numeracy and health and wellbeing.' [47] The SAC has three main strands: Challenge Authorities, Schools Programme, Pupil Equity Funding.

The Attainment Scotland Fund was set up to administer £750 million of targeted financial support for the Challenge strands, from 2016 to 2021, to raise the attainment of pupils in local authorities, and particular schools outside these, with the highest concentrations of deprivation. There are nine 'Challenge Authorities': Glasgow, Dundee, Inverclyde, West Dunbartonshire, North Ayrshire, Clackmannanshire, North Lanarkshire, East Ayrshire and Renfrewshire. The initiative focussed on primary schools initially but has since been extended to 133 secondary schools with 20% or more of their pupils coming from SIMD 1 and 2, the most deprived 40% of households. Funding of £11.5 million has been earmarked for projects designed to close the attainment gap. In the main, these funds have been allocated to schools in the nine Challenge areas. [48] The funding is to support secondary schools to design their own, context-sensitive, project interventions to improve the literacy, numeracy, health and wellbeing of their pupils living in areas of deprivation.

The Pupil Equity Fund (PEF) is provided as part of the Attainment Scotland Fund with £120 million to be distributed between 2017 and 2018. These funds are allocated directly to schools, to be spent at the discretion of head-teachers on additional staffing for resources specifically to target the poverty-related attainment gap. PEF funding has been allocated to 95% of schools for pupils in P1-S3 known to be eligible for FSM. [49] Schools receive £1,200 for each FSM eligible P1-S3 pupil (Scottish Government, 2018).

[47] https://education.gov.scot/what-we-do/delivering-the-scottish-attainment-challenge/About%20the%20Scottish%20Attainment%20Challenge, accessed 21/07/2018.
[48] Funding has also been allocated to 28 secondary schools outside the nine Challenge Authorities (https://education.gov.scot/what-we-do/delivering-the-scottish-attainment-challenge/Schools%20programme%20(secondary), accessed 21/07/18.
[49] https://beta.gov.scot/policies/schools/pupil-attainment/, accessed 21/07/18.

The recent interim evaluation report for the first two years of the Attainment Scotland Fund suggests that there was some evidence to indicate that the attainment gap was smaller in Challenge Authorities compared to non-challenge authorities and that 'pupils from areas of greater deprivation performed better in Challenge Authorities than in other areas of Scotland.' (*ibid*). It should be noted, however, that, as highlighted by Raffe et al. (2006, documented in Chapter One), educational inequalities tend to be stubbornly stable over time; any change that takes place tends to be small and occurs over decades reflecting wider socio-economic change rather than just educational reforms.

In 2013/14, SIMD 5 individuals, those from least deprived (most affluent) 20% of households were almost three times as likely as those from SIMD 1, the most deprived 20% of households, to leave school with three Highers (Commission on Widening Access, 2015). UCAS statistics indicate that individuals from SIMD 5 are more than four times as likely to go to university compared to those from SIMD 1 and that the position is considerably worse when considering the most selective institutions (Commission on Widening Access, 2016). The Scottish Government has committed itself to equality of access to HE by 2030 in terms of the representativeness of the student population, whereby 20% of HE entrants are to be from SIMD 1 households. In order to achieve this, in its final report, the Commission on Widening Access (*ibid*), has recommended that by 2019, there should be *access thresholds* for all degree programmes in Scottish universities against which the those from the most deprived backgrounds should be assessed for HE entry. The proposal is for a two-tier entry system with access thresholds set separately from standard entrance requirements to reflect 'the minimum academic standard and subject knowledge necessary to successfully complete a degree programme.' (*ibid*, p13). This recommendation was fully accepted by the Scottish Government.

Given the imminent introduction of access thresholds for degree entry, and the importance of subject choice as outlined in Chapter One, this chapter examines attainment in Facilitating Subjects for the years 2002 to 2009. The aim is to assess the extent of any attainment gap once, arguably better, facilitating subject choices have been made and consider the justification for a two-tier HE entry system. The next section reviews the broad literature examining attainment in Scotland and the UK. This is followed by an explanation of how the different qualifications' data were standardised into pass grades for analysis and the methodology used. The results and analysis are then presented, followed by concluding comments.

2.3 The Attainment Literature

2.3.1 Trends in Educational Attainment and Participation in Scotland

Analysis of trends in attainment exploring the impact of different regional education policies indicates that social-class inequality at age 16 narrowed in Scotland (but not England) from the mid-1980s to the mid 2000s (Croxford & Raffe, 2007). It is suggested that this observed narrowing may have been attributable to curriculum and assessment reforms associated with the introduction of SG qualifications. SG qualifications were designed to certificate all levels of attainment to increase the proportion of young people leaving school with formal qualifications. [50] In earlier work, Raffe et al. (2006) reported that within each Scottish cohort, the class gap was wider at 18 than at 16 and that, by the late 1990s, inequalities at 18 were substantially wider in Scotland than England. They maintained that social class differences in HE entry are largely attributable to class differences in achieving the required entry qualifications and that these inequalities in entry to degree courses are wider than those across HE as a whole. This finding is in keeping with similar analysis for England and Wales where inequality in access to HE rose throughout the 1980s and 1990s despite the increased post-16 participation and attainment of pupils from low income backgrounds (Blanden et al., 2003).

Time Series Analysis of the Scottish School Leavers' Survey (SSLS) by Croxford (2009) supports the finding of Raffe et al. (2006) that increasing levels of attainment and participation at 16 have pushed the critical period for educational inequalities up to age 18. The study revealed upward trends in both participation and attainment at age 16 and 18 over the period 1985-2005. These trends were associated with an underlying polarisation between those with high and low SG performances; participation in Further or Higher education at age 18 increased proportionately more among those in higher attainment SG bands compared with those in lower attainment SG bands. There appeared to be a marked increase in 'staying on' after 16 among lower attaining groups whereas,

[50] The minimum achievement target was changed from 5 'O' grades at grade C or above to 5 SGs at SCQF Level 3 (that is *Foundation* level SG 5/6) or above. The SG equivalent to a C at 'O' grade was *Credit* level SG 2.

historically, staying on has been linked to the level of S4 attainment with higher attainment leading to a greater likelihood of staying on. SG attainment was also seen to be a predictor of S5 Higher outcomes; students with middle and low SG attainment had poorer average pass rates at Higher than those with high SG attainment. Social class was found to be a major source of inequality in attainment with class differences being wider than gender differences. The gap in SG average point scores between pupils from a Managerial/Professional class background and those from a working class background narrowed from 16 to 10 points (32 and 16 compared to 44 and 34, respectively). Social class inequality in the attainment of UCAS tariff points for university entry at 18 was very pronounced with individuals from all other social classes having substantially lower average attainment than those from Managerial/Professional class backgrounds. This was despite working-class attainment rising significantly between 2001 and 2005 with the Higher Still reforms. Females were seen to have higher attainment at both 16 and 18 and this gender gap increased for both age groups over time. School intake characteristics, particularly the proportion of students from managerial/professional families, were seen to have additional effects on attainment with these reducing over time at age 16 but remaining the same at age 18. Individuals attending schools with a high percentage of pupils from Managerial/Professional backgrounds tended to have higher average UCAS tariff scores; the score of those attending schools with a high percentage of working class students did not differ from the average. Schools in remote localities were seen to have higher age 16 attainment than schools in cities, urban areas and other accessible areas. Schools in the four big cities, however, consistently had higher age 18 average attainment than other schools.

It is young people from low-income households that tend to have low attainment and are less likely to participate in post-compulsory education (Croxford et al., 2004). While a substantial proportion of low achievers continue at school beyond the end of compulsory education, the majority of those who stay on tend to come from more advantaged backgrounds. ‘Stayers’ are more likely to obtain a formal qualification by the time they reach 22-23 years-old compared with low attainers who leave early, 63% as opposed to 14% (Howieson & Iannelli, 2008). Research in other countries also suggests the transition to post-compulsory education is still a critical threshold for the transmission of educational inequality; individuals are more likely to be unemployed or employed in unskilled jobs if they do not proceed beyond compulsory education (see, for instance, Bernardi, 2012).

In Scotland, prior to the introduction of CfE[51], improved opportunities for access do not appear to have translated into improved attainment for mid-low attainers despite the implementation of the *Higher Still* framework policy initiative (Raffe et al., 2005, 2007). The introduction of this new national qualifications' framework in 1999-2000 was followed by a substantial overall increase in participation and attainment. The steepest increase in participation, however, was seen among those with low SG attainment whilst the steepest increase in attainment was among those with high SG attainment (Croxford et al., 2004). Pirrie and Hockings (2012) surmise that, in general, targeted initiatives seem to work in terms of closing gaps (at least) in the short-term but universal roll out does not as the middle classes tend to benefit disproportionately (Hills et al., 2010). It is clear that Mare's distinction between the distribution and allocation of education applies to Scotland (Mare, 1981). As Paterson et al. (2011) conclude, widening access is not sufficient to reduce educational inequalities; the ability to take advantage of opportunities needs to be widened.

2.3.2 Patterns of Attainment

Patterns of attainment in terms of reoccurring associations between pupil characteristics in terms of gender and socio-economic background, are examined here for both Scotland and the wider UK since the education systems are more similar than they are different and are politically and functionally interdependent (Raffe, 2000). Attainment is often measured and analysed in terms of some chosen benchmark number of awards at a particular qualification level for the purpose of setting policy targets and the academic study of such. This pragmatic, but at the same time rather simplistic, approach may cause more nuanced patterns of attainment to be missed. The debate as to which students are driving average attainment statistics and what lies beneath is ongoing.

[51] An early stage evaluation of the impact of CfE on attainment, analysing data at the LA level, indicated that whilst the percentage of school leavers with Highers (the crucial qualifications for HE entry) increased across all levels of deprivation at the LA level, this increase was most noticeable in more deprived LAs (Shapira & Priestley, 2018).

The age-16 attainment of both females and males has increased substantially since 1965, but the rise in female attainment has been greater than that for males with lower average attainment for the latter being seen since 1975 (Tinklin et al., 2001). This gender gap – the relative underachievement of boys - has received much attention in educational discourse for the last 40 years or so with conflicting evidence as to whether it is widening over time (Burgess et al, 2003). The underachievement of boys at the highest grades was found to drive the gender gap by Gorard et al. (1999) with no uniform distribution of a gender gap in evidence across the attainment range. Boaler et al. (2000), on the other hand, identify underachievement by girls at the top end of the ability distribution and suggest that this may be linked to the more pressurised, high expectation environment found in top-stream classes.

Examining gender differences in pupil performance in Scottish schools, Tinklin et al. (2001) observed from their review of the literature that all explanations lie on a nature-nurture continuum with the overtly biological tending not to be favoured as this might justify discrimination. Differences between females and males in attitudes, confidence, behaviour and classroom interaction were identified and variously attributed to: peer pressure, different approaches to assessment and curricular tasks, teaching and learning processes, perception and influence of post-school opportunities. Girls were found to lack confidence in their own abilities, while boys were more confident, more likely to contribute to classroom discussions but more susceptible to peer pressures. Girls were seen to perform better than boys across all school levels - pre-school, primary and secondary. In terms of formal qualifications, analysis of relative odds ratios revealed differences in uptake and attainment in different subjects. [52] Compared to boys, girls were more likely to gain awards in almost all their SG subjects and passes at A -C in almost all their Higher subjects, with females more likely to achieve A grades. In some Higher subjects, however, males were found to be more likely to achieve A grades; these were: modern languages, Mathematics, Biology, Chemistry, Economics, Accounting and Finance. No systematic gender differences between schools were found. While gender differences in educational achievement were seen across all social classes, the average

[52] The authors variously used SQA, Scottish School Leavers' Survey (SSLS) and published official statistical bulletin data.

effect of gender on attainment was seen to be much smaller than that for social class. Tinklin et al conclude that average attainment statistics conceal differences within the two genders as females and males are not homogenous groups; there are high attaining males and low achieving (or failing) females. They assert that the differences between high and low attainers (of both sexes) are greater than those between females and males.

In a more recent review of the literature on gender inequalities in schools (Forde et al., 2006) surmise that girls and boys do appear to relate differently to schooling and learning with girls finding it easier to succeed in school settings. Much of the evidence suggests gender to be a socio-cultural construct with the formation of gender identities starting with early family experiences and continuing throughout school. Children's identities are observed to be multi-faceted, evolving though interaction and negotiation in different social and cultural contexts, including school. The OECD (2015) reported that, in general, across the world, boys work less hard in school than girls. Boys were found to be: 8% more likely to view school as a waste of time, ten times more likely to play online, collaborative games (20% : 2%), less likely to read for enjoyment (60% : 77%) and spend one hour less on homework per week (4.2 compared to 5.5 hours for girls).

Gender differentiated patterns of attainment between ages 14 and 16 in English state secondary schools are investigated by Burgess et al. (2003). They examined matched exam results, from national Department for Education and Skills (DfES) data, for more than half-a-million children in 3,000 plus schools (the entire cohort), who sat Key Stage 3 (age 14) tests in 1999 and GCSEs (age 16 formal qualifications) in 2001. The gender gap, the phenomenon of the relative underachievement of boys, is investigated both in the aggregate and at subject level focussing on English, Maths and Science; as study of these tends to be compulsory and, therefore, less prone to selection issues. Three measures of attainment were used: the percentage of boys and girls gaining at least five GCSEs awards at grades A*-C (the widely used policy target measure in England), total GCSE points, value added between 14 and 16. The effects of school and performance, gender mix, admissions' policy, pupil eligibility for FSM and LA influence are investigated. The gender gap was found to be constant across all data/model permutations; across both the ability and attainment distributions and for the dependent variable whether measured by GCSE points' scores or value added. School quality, whether a school performed well or poorly, or was effective or ineffective and other observable school characteristics were seen to have no effect. The primary driver was seen to be performance differentials in English scores. Wide gender differences were in evidence across the range of both the prior attainment and GCSE outcome distribution but were particularly noticeable at the lower quartile. For Maths and Science, in general, given prior attainment levels, median males and females achieved the same GCSE score. Focussing on the top decile of children for prior attainment, however, revealed that the Maths and Science GCSE points scores for boys were higher than those for girls.

School level gender gaps, defined as the mean difference between girls' and boys' total GCSE points, were regressed against a set of school characteristics. This revealed a negative relationship between the gender gap and both FSM eligibility and the proportion of boys in school cohorts. Lower poverty levels were seen to be associated with increased gender-differences in performance/attainment while an increased proportion of boys within a school cohort reduced the gender gap. Burgess et al. (*ibid*) speculate that this latter finding could be the effect of either boys performing relatively better, or girls relatively worse, in cohorts containing higher proportions of boys. They emphasize that the gender gap was not seen to be influenced in any great way by observable school characteristics and that the *ability* and *poverty gaps* (measured by prior attainment and FSM eligibility respectively) were found to be far larger. They posit that the source of the gap would appear to be generic (societal or physiological in nature) rather than determined by the behaviour of schools and teachers, and that this should be taken into consideration in the formation of education policy. They speculate further as to whether, whilst there is a clear difference between subjects at present, girls might come to outperform boys in the traditionally male areas of Maths and Science too, if some underlying, slow socialisation process is taking place. Alternatively, it is suggested that the different cognitive demands and processes required by different subjects indicate that the gender gap may be genetic, 'rooted in different rates of cognitive maturation between boys and girls, that itself happens at varying rates for different cognitive processes' (*ibid*, p12).

Reviewing gender, race and class factors, Hastings (2006) raises issues of social capital and cultural attitudes with class being argued to cut across gender and race as the most important determinant of educational attainment. Research on intergenerational mobility in the UK in 1980s and 1990s (Blanden et al., 2003), suggests that there has been a sharp rise in tertiary educational inequality. Parental income was seen to matter less for the staying on at school decision in the 1990s but the rapid expansion of HE, coupled with regressive changes in student financial support, was seen to disproportionately benefit those from more affluent backgrounds. In further work, the authors suggest that the increased sensitivity of education to parental income that occurred in the 1990s has led to a sharp fall in intergenerational mobility in the UK (Blanden et al., 2004).

Using several large-scale surveys, [53] Goodman and Gregg (2010) investigated the potential role of aspirations, attitudes and behaviours in explaining the attainment gap between economically disadvantaged and more affluent children. They found that, while the growth in the attainment gap slows in secondary school compared to primary, it is very large by the time children come to sit GCSE exams. Measuring relative (dis)advantage by quintiles based on parental socio-economic position, 75% of the top quintile were seen to gain five GCSE passes at grades A*-C compared to only 21% from the bottom quintile; a 54% gap. The aspirations, attitudes and behaviours of both teenagers and their parents were seen to influence the GCSE attainment gap. Individuals were more likely to have better GCSE results if, for instance, their parents thought that they would go to university and devoted material resources to education. Self-belief and aspirations were found to be important, with individuals more likely to do well if they believed that events resulted primarily from their own actions and that they would apply and get into university. [54] Additionally, children's attainment was seen to have a clear *heritability* element with nearly 20% of the test score gaps between the most and least advantaged children being "explained by an apparent 'direct' link between the childhood cognitive ability of parents and that of their children" (*ibid*, p7).

More recently, Andrews et al. (2017) focus on the attainment gap between disadvantaged 16-year olds (defined by FSM eligibility) and their peers in England in terms of individuals'

[53] Millennium Cohort Survery (MCS), Avon Longitudinal Study of Parents and Children (ALSPAC), Longitudinal Study of Young People in England (LSYPE), British Cohort Survey (BCS).

[54] Other attitudinal / behavioural factors that were seen to be important were: shared family meal times and outings, less parent-child quarrels, the avoidance of risky behaviours by teenagers (smoking, cannabis use, anti-social behaviour, truancy, suspension, exclusion) and not being subject to bullying (*ibid*).

relative position on the attainment distribution using National Pupil Database (NPD) data. All pupils were ranked from highest to lowest according to their attainment. Mean ranks were calculated for the disadvantaged group and the non-disadvantaged (non-FSM) group. The attainment gap was measured as the difference between these ranks and subsequently converted into months of progress. It was reported that the attainment gap between disadvantaged 16-year olds and their peers had narrowed very slightly between 2007 and 2016, by the equivalent of 3 months of learning. On average, however, disadvantaged pupils were seen to fall behind their better off counterparts by two months for each year of secondary school.

The relative importance assigned to gender and poverty gaps is reversed by Hillman and Robinson (2016) who claim that poor educational attainment by those from disadvantaged socio-economic backgrounds and particular ethnic groups can only be addressed properly if gender inequalities are dealt with. Concerned with male underachievement in terms of relatively lower HE entry, higher drop-out rates and lower degree performance, they argue that while men underperform in general, it is poor white men who have the worst attainment. It is argued that tackling the underperformance of young men prior to age 18 is essential if poor male HE performance is to be addressed and that broad socio-economic background measures such as SIMD quintiles may hide more a more complex picture. They cite recent UCAS analysis [55] that reports a 28% (age 18) HE entry rate for state school pupils from POLAR quintile 3 areas. [56] Within quintile analysis by different population characteristics reveal a 9% entry rate for white men with FSM entitlement compared to, for instance a 44% entry rate for non-FSM eligible Asian women. This variation within quintile 3 is larger than that between the most advantaged and most disadvantaged quintiles 1 and 5.

In their work Gayle and Playford (2014) and Gayle et al. (2016), examine age-16 attainment in terms of broader subject areas. Using a latent class model approach with

---

[55] UCAS, End of Cycle Report 2015, p.130

[56] POLAR is the Participation of Local Areas classification, a UK-wide area-based measure that groups geographical areas into quintiles according to the proportion of young people living in them who participate in HE by the age of 19. This is known as the "young participation rate". Amongst other things, the POLAR classification is used to distribute HEFCE's student opportunity allocation to HE institutions and to monitor local and national HE participation patterns of young people.
From http:// www.hefce.ac.uk/pubs/year/2014/201401/.

YCS data [57] for England and Wales (Gayle and Playford, 2014), they identified four distinct educational groups. At opposite ends of the attainment spectrum were Group 1, high achievers with good GCSE attainment across all subject areas and Group 4, low achievers with poor GCSE attainment across all subject areas. Groups 2 and 3 were characterised by similar, middling levels of attainment in terms of the number of their awards and associated points but were seen to have distinct aptitudes in different subject areas. Group 2 had good attainment in Science while Group 3 tended to perform better in *Arts* (Humanities) subjects and performed poorly in Science and Maths. Group membership was found to be stratified according to socio-economic background in terms of parental occupation [58] with this being the most important predictor. Pupils from more advantaged backgrounds (where parents had professional or managerial occupations) were more likely to be found in Group 1 and those from less advantaged backgrounds (with parents in semi-routine or routine occupations) were more likely to be in the low attaining Group 4. The middle ground of Groups 2 and 3 tended to be occupied by children with parents in intermediate occupations with no significant difference in socio-economic background effect between the two groups. Additionally, in these middle two groups, a significant gender effect was evident; being male made membership of the Science Group 2 more likely but membership of Arts Group 3 less likely.

The same latent class modelling approach was used with SQA administrative data, held as part of the Scottish Longitudinal Study (SLS), by Gayle et al. (2016), to analyse pupils' SG subject area outcomes for Scotland for the years 2007 to 2011. Similar to the analyses for England and Wales, four main educational groups were identified. Again, at either end of the achievement spectrum, one group had very good attainment with individuals in this group tending to come from more advantaged backgrounds, whilst another group had very poor attainment and, generally, were seen to come from more disadvantaged backgrounds. Two middle attaining groups, with similar overall outcomes in terms of the number of awards and associated points, were characterised by distinct differences in terms of subject area. One of these groups was more likely to obtain an SG Credit pass in English but less likely to achieve this in Maths and Sciences. The other middle attaining

[57] Data for YCS Cohort 6, surveyed in 1992, were used.
[58] As measured by National Statistics Socio-economic Classification (NS-SEC).

group were unlikely to secure credit passes in either English or Maths but more likely to do so in Sciences. There were similar gender differences with boys being seen to do less well overall; they were more likely to be in the poor attainment group rather than the high attainment group. In terms of the middle attaining groups, boys were more likely to be in the science group rather than the non-science group. It should be noted, however, that this work did not include Intermediate 2 qualifications that were offered in tandem with SGs at a similar level in the years examined and so does not comprehensively cover age-16 formal attainment in Scotland.

Motivated by concerns that the underachievement of boys maybe be class- and race-based, with working-class boys and those from certain ethnic minorities [59] performing significantly less well, Connolly (2006) investigated the impact of social class and ethnicity on gender differences in GCSE attainment. YCS data for three successive cohorts (1997, 1999, 2001) were used to assess whether there might be interaction effects between class and gender, and ethnicity and gender, whereby particular combinations of these either reduce or increase gender differences in educational attainment. He reported that, for all three cohorts, class and ethnicity have a far greater impact on age-16 attainment than gender. For the 2001 cohort, using logistic regression analysis, girls were seen to be 62% more likely than boys to gain five or more A*-C GCSE passes. Individuals with parents from higher professional occupations and lower professional occupations were eight and four times more likely to achieve five or more A*-C GCSE passes compared to those with parents in routine occupations. Boys from the highest parental occupational backgrounds were found to be two to three times more likely to obtain five or more GCSE grades A*–C compared to girls from the lowest parental occupational backgrounds. With respect to ethnicity, Chinese respondents were seen to be seven times more likely to obtain five or more GCSE grades A*–C when compared to Black respondents. No systematic variation in the size of gender differences across social classes or ethnic groups was found, with these appearing to be relatively stable and constant. Connolly concludes gender effects would appear to be independent of social class and ethnicity effects.

[59] African Caribbean, Bangladeshi and Pakistani boys in particular.

Hupkau et al. (2016) argue that to address low social mobility and underachievement, it is necessary to examine progression routes for the substantial numbers of individuals who opt for vocational qualifications at age 16; as the labour market value of many of these can be quite variable whilst at the same time they may not provide a route to higher levels of education. It is suggested that this is a much-neglected area of research due to the complexity and diversity of vocational qualifications 'with thousands of available courses varying widely in length, level, degree of difficulty and specialisation' (*ibid*, p1). Moreover, gendered post-school vocational education destinations are still much in evidence and tend to disadvantage females (Forde et al, 2006). Hupkau et al. (2016) use linked administrative data to track the educational choices made over a four-year period, from age 16 to age 20, by all students in England who sat GCSE examinations in 2010. They group multifarious post-16 qualifications into several broad categories and examine the probability of achieving various outcomes at age 20 given a student's chosen educational path at age 17. The various outcomes examined were those that are known to have positive labour market returns (see for instance: Blundell et al. 2005, Dearden et al. 2002, McIntosh, 2006 (for England), Gasteen and Houston, 2007 (for Scotland). These were: staying on in education up to age 18, securing an upper secondary (Level 3) qualification, commencing an undergraduate degree, attending a Russell Group university, commencing some other form of tertiary (Level 4 and above) education, commencing an apprenticeship. Students were classified according to their highest level of learning and age 17 main educational activity; this proved to be relatively simple for vocational qualifications equivalent to 'A' Levels [60] but was not possible for lower level qualifications as the diversity was too great.

Embarking on 'A' level courses or vocational equivalents were found to be equally strong predictors of the likelihood of staying in education up to the age of 18 and achieving a Level 3 qualification before the age of 20. As might be expected, 'A' Levels were seen to be very much the dominant pathway for access to HE and the prestigious Russell Group universities in particular. Apprenticeships were classified as *intermediate* (Level 2) or *advanced* (Level 3) [61] with 20% of the cohort observed as being on such at some point

[60] Level Three in the English and Welsh Education System.
[61] As Hupkau et al. (2016) point out, there will enormous variation within these categories with some apprenticeships being more demanding and prestigious than others.

between the ages of 18 and 20. Individuals accessing Level 2 apprenticeships tended to have lower than (the cohort) average GCSE attainment. Those accessing advanced apprenticeships (some 40% of the total on apprenticeships) were seen to have slightly higher than average achievement but their profiles differed markedly from those who took 'A' levels and entered university. The latter generally had much higher attainment and were less likely to come from a low-income background; once again confirming the established pattern of *who goes where* based on socio-economic factors.

Poor GCSE attainment can impact negatively on young people in early adulthood and beyond in terms of less favourable labour market outcomes (Murray, 2011, Jones et al., 2003) with lower earnings' potential (Dearden et al., 2004) and the risk of *wage scarring* effects occurring as a result of youth unemployment (Gregg and Tominey, 2005). Poor secondary school achievement has been found to be a more important explanation of low HE participation among individuals from poorer socio-economic backgrounds than any potential entry point barriers (Chowdry et al., 2013). It would appear that this is not something that can easily be rectified. Hupkau et al. (2016) found no clear trajectory to higher levels of education for individuals with poor age-16 GCSE attainment who, as a result, were studying for Level 2 qualifcations at age 17. Of those pursuing Level 2 qualifications, less than half (44%) were seen to achieve a Level 3 qualification by age 20. For individuals studying for Level 1 or below qualifications at age 17, only 16% went on to obtain a Level 3 qualification by age 20. Worryingly, many of those observed studying at low qualification levels were seen to do so for several years despite the usually short duration and part-time availability of such; 10,000 students in this single cohort were seen to be studying for low level qualifications for four consecutive years. This research raises the uncomfortable question as to the effectiveness of the provision of some FE vocational courses as a *second chance* for substantial numbers of individuals who leave school with low attainment.

Qualifications in facilitating subjects enhance individuals' chances of entering both more prestigious universities and degree programmes that attract higher labour market premia. Class and gender differences are evident in the uptake of the *number* of facilitating subjects at Higher, the crucial level for HE entry in Scotland, as shown in Chapter One. Uptake alone, of course, is not sufficient as grades are of paramount importance in the competition for places at more prestigious universities. Attainment levels in facilitating subjects are now examined; the methodology used is outlined below followed by the results and analysis.

2.4 Data and Methodology

The weakness of measuring attainment in terms of a benchmark number of awards (or associated points) is that a given pass grade in, for instance, Maths is regarded as equivalent to the same pass grade in Business Studies or Physical Education. This is unsatisfactory because, as discussed in Chapter One, some (facilitating) subjects have more currency than others when applying for tertiary education courses. It is also unsatisfactory to measure attainment at the critical S5 and S6 level simply in terms of a pass as the grades achieved can be very important for HE entry and course access. With this in mind, attainment in facilitating subjects by level of pass is investigated below.

The individual dependent variables were created by recoding students' UCAS points for each facilitating subject, at each qualification level, into four ordinal performance levels as shown in Table 2.2. As each dependent variable has four outcomes, there are three thresholds that partition the range of the exam/coursework marks into the grade bands. If an individual's mark lies above the first threshold but below the second, this implies that they have passed the subject at the lowest passing grade. A mark above the second or

Table 2.2

STEM Subject Performance Levels for Ordered Logit Dependent Variables [62]

SG UCAS points / (grade)	Subject Level Band	Int2 UCAS points / (grade)	Subject Level Band	H UCAS points / (grade)	Subject Level Band	AH UCAS points / (grade)	Subject Level Band
38 (1)	Top	42 (A)	Top	72 (A)	Top	120 (A)	Top
28 (2)	Middle	35 (B)	Top	60 (B)	Middle	100 (B)	Middle
22 (3)	Middle	28 (C)	Middle	48 (C)	Bottom	80 (C)	Bottom
16 (4)	Bottom	24 (D)	Middle	42 (D)	Fail	70 (D)	Fail
11 (5)	Bottom						
8 (6)	Fail						
3 (7)	Fail						

[62] Derived from the Unified Points Score Scale, http://www.gov.scot/Publications/2009/03/09154229/3 accessed 14/06/2018.

third thresholds implies a pass at the middle or higher grade respectively, while a mark below the first threshold implies a poor/failing performance.

As discussed in the Introduction, subject grade boundaries are set annually by SQA which can result in these changing across cohort years, reflecting both within cohort attainment and the judged degree of difficulty of the assessment. Clearly, graded passes in qualifications are *ordered* outcomes as the higher the grade, the higher the level of attainment in a specific subject. This suggests that it would be appropriate to model subject attainment outcomes using an ordinal regression model where individuals are observed in one of a number of ordered categories. In keeping with the analysis in Chapter One, ordered logit estimation was preferred to ordered probit as odds ratios can be reported directly and, in practice, there is very little difference between the two but the latter is more mathematically complex (Gujarati, 2015). In addition to regressor coefficients, ordered logit models estimate threshold parameters or *cut points* at which an individual will move from one category into another. The categories differ only in these intercepts, the slope coefficients for the independent variables are the same for each category; this is the assumption of *proportional odds*. Essentially, one model is estimated for all categories/levels of the dependent variable with the impact of any of the regressors assumed to be the same in each category giving a series of parallel regression lines. The assumption of proportional odds was tested using Long and Freese's *omodel* routine STATA add-in (Long and Freese, 2006). The routine tests the null hypothesis that the beta coefficients for each regressor, at each level, are the same using a chi-square test. The assumption of proportional odds, the null hypothesis, is rejected if the chi-square test statistic is significant. In the event, with just two exceptions (Advanced Higher Physics and French), all the ordered logit models failed the omodel test for proportional odds, see Appendix B, Table B1. It may be that the annual setting of grade boundaries by SQA undermines the assumption of proportional odds as these boundaries may shift position across years. In the event, multinomial logit models were estimated instead.

In multinomial logit models (MLMs), there are more than two outcomes for the dependent variable but these have no natural order, that is the alternative outcomes are *nominal* rather than ordinal. MLMs are appropriate where the data are *chooser-specific*, i.e. where an individual has to choose between several alternatives and the regressors vary across individuals (Gujarati, 2015).[63] An alternative approach would have been to use generalised ordered logits. MLMs were considered to be appropriate as the priority was to capture the potentially varying effects of gender and socio-economic background across the different outcomes rather than maintain a strict order of pass grades.

Generalising the bivariate logit model as outlined in Chapter One, produces the following MLM, where X and β are vectors of the independent variables and their respective coefficients, and α is the constant term.

$$p_{ij} = \frac{e^{a_j+\beta_j X_i}}{\sum_{j=1}^{4} e^{a_j+\beta_j X_i}}$$

This expression gives individual i's probability (p_{ij}) of choosing option one (j=1) as opposed to the three other options where the sum of the probabilities of these mutually exclusive outcomes adds up to 1. The subscripts on both the intercept *and* slope coefficients indicate that the values of these can vary between the different outcomes (as opposed to the proportional odds' assumption of the ordered logit model where the beta values are identical). Essentially, the model allows for the significance and/or impact of the explanatory variables to differ between outcomes by estimating a *separate* regression for each outcome. As the sum of the probabilities for the four outcomes must sum to 1, they cannot all be estimated independently; if three of the probabilities are calculated, the fourth must be determined automatically to ensure that they sum to 1. The general practice is to choose one outcome as the reference category and set its coefficient values to zero (*ibid*). Designating Fail, Low Pass, Middle Pass and High Pass as outcomes 1, 2, 3 and 4 respectively and choosing Fail as the reference category gives the following probabilities for the four attainment outcomes:

$$p_{i1} = \frac{1}{1 + e^{a_2+\beta_2 X_i} + e^{a_3+\beta_3 X_i} + e^{a_4+\beta_4 X_i}}$$

[63] Multinomial probit models (MPMs) would also be appropriate but, as with ordered probits, are more mathematically complex and so MLMs are preferred here, following Gujarati (*ibid*).

$$p_{i2} = \frac{e^{a_2+\beta_2 X_i}}{1 + e^{a_2+\beta_2 X_i} + e^{a_3+\beta_3 X_i} + e^{a_4+\beta_4 X_i}}$$

$$p_{i3} = \frac{e^{a_3+\beta_3 X_i}}{1 + e^{a_2+\beta_2 X_i} + e^{a_3+\beta_3 X_i} + e^{a_4+\beta_4 X_i}}$$

$$p_{i4} = \frac{e^{a_4+\beta_4 X_i}}{1 + e^{a_2+\beta_2 X_i} + e^{a_3+\beta_3 X_i} + e^{a_4+\beta_4 X_i}}$$

Taking the log of these odds ratios gives the individual logit regression models that are linear functions of the explanatory variables, viz:

$$ln\left(\frac{p_{i2}}{p_{i1}}\right) = a_2 + \beta_2 X_i$$

$$ln\left(\frac{p_{i3}}{p_{i1}}\right) = a_3 + \beta_3 X_i$$

$$ln\left(\frac{p_{i4}}{p_{i1}}\right) = a_4 + \beta_4 X_i$$

$$p_{i1} = 1 - p_{i2} - p_{i3} - p_{i4}$$

The logits are estimated simultaneously using Maximum Likelihood and give the change in the logarithmic chance of preferring one of the options (pass grades) to the reference category (Fail) for a one unit increase in an explanatory variable, holding all the other variables constant. The change in the logarithmic chance can be converted to an odds ratio by taking the anti-log. Generalising, as explained in Chapter One, the anti-log $e^{\beta 1}$ gives the odds ratio for the independent variable X_1; that is, the increase in the odds of an outcome associated with a one unit increase in X_1. Fail (option 1) was chosen as the base outcome and STATA's Relative Risk Ratio (rrr) was specified to directly provide odds ratios. In the analysis below, the odds indicate by how much one of the pass grades is a more likely outcome than a Fail, the comparison category.

It is not possible to safely compare log-odds ratios or odds ratios from logistic regressions across groups within a population or sample, even when the same independent variables are used in the models, as unobserved heterogeneity can vary across groups causing log-odds to be rescaled (Mood, 2010). For this reason, average marginal effects are reported.

The average marginal effect (AME) is given by:

$$\frac{1}{n}\sum_{i=1}^{n}\beta_{x_1}f(\beta x_i)$$

where β_{x1} is the estimated log-odds ratio for independent variable x_1, β_{xi} is the value of the logit (i.e. the linear combination of the values of the x variables and their estimated coefficients) for the i-th observation, and f(xi) is the logistic probability distribution function (PDF) with respect to βx_i. The AME measures the average effect of x_1 on P(y=1) - the probability that y equals one - for the population under consideration "… by taking the logistic PDF at each observation's estimated logit, multiplying this by the coefficient for x_1, and averaging this product over all observations." (*ibid*, p75). This is the average change in probability when x_1 increases by one unit. As logistic regression models are non-linear, the effect of a change in x_1 will differ from individual to individual; the AME calculates the effect for each individual and then calculates the average for the population/sample. AMEs can be compared across models and groups as they are not affected by any unobserved heterogeneity that is unrelated to a model's independent variables (*ibid*, p78).

2.4.1 Model Specifications

MLMs were estimated for 2 broad groups of facilitating subjects at all qualification levels. The groups were: Humanities and Languages comprising English, Geography, History, Modern Studies, French, German, Spanish, and Maths and Science comprising Maths, Biology, Chemistry, Physics. These subjects were analysed on the basis that they might be considered to be the *main stream* facilitating subjects as identified by the Russell Group of universities. A number of subjects included in the Chapter One analysis, in an attempt to keep the scope of what might be considered a facilitating subject relatively broad, were excluded on the basis of their narrow provision/uptake. These were mainly languages (Russian, Italian, Gaelic, Urdu, Latin and Classical Studies/Ancient Greek). SGI level general science was also excluded since, although provision and uptake was widespread, this tends to be taken by relatively weaker pupils or those with no interest in studying a specific named science.

For each subject, at each qualification level, as indicated in the above discussion, the dependent variable had four possible outcomes: 1 - Fail, 2 - Low pass, 3 - middle pass and 4 – high pass, with Fail being specified as the reference category. The baseline generic model is given by the following expression; this is the likelihood that individual *i* will obtain a low pass (p_{i2}) as opposed to a fail (p_{i1}) in a facilitating subject at SGI level.

$$
\begin{aligned}
ln\left(\frac{p_{i2}}{p_{i1}}\right) = a_2 + \beta_{2|1} Sex_i + \sum_{j=2}^{6} \beta_{2|j} SIMD1-5_i + \beta_{2|7} SGI\ cohort\ size_i \\
+ \beta_{2|8}\%\ pupils\ SIMD\ 1\ \&\ 2_i \\
+ \sum_{k=9}^{14} \beta_{2|k} UrbanRural_i + \sum_{l=15}^{46} \beta_{2|l} LA + \sum_{m=47}^{54} \beta_{2|m} Year + \varepsilon_i
\end{aligned}
$$

Where:

- Sex (Male=1)
- SIMD 1–5: individuals' household Scottish Index of Multiple Deprivation 2009 quintile
- SGI cohort size: school size as measured by Standard Grade/Intermediate 2 cohort size
- % pupils SIMD 1 & 2: schools' socio-economic composition given by the percentage of pupils from SIMD quintiles 1 and 2, the most deprived 40% of households
- Urban/Rural: dummy variables for urban/rural location and degree of remoteness
- LA: local authority dummy variable for individuals' school
- Year: dummy variables for years 2002-2009

- The Higher and Advanced Higher models contained measures of individuals' previous absolute and relative (within school) attainment and the size of the particular qualification cohort viz:
- SGI UCAS points: total Standard Grade/Intermediate 2 UCAS points at age-16 (Highers' models)
- Higher UCAS points: total Highers' UCAS points at age-17 (Advanced Highers' models)
- School quartiles 1-4: individuals' SGI UCAS points' school quartile (Highers' and Advanced Highers' models)
- Higher cohort size: school size as measured by Higher cohort size (Highers' models)
- AH cohort size: school size as measured by Advanced Higher cohort size (Advanced Highers' models)

Previous absolute attainment was given by individuals' total SGI UCAS points for the Highers' models and their total Higher UCAS points for the Advanced Highers' models. Relative attainment, for both sets of models, was given by individuals' School quartile based on their SGI UCAS points.

Summary statistics for the data used in MLM models are provided in Appendix B, Table B2. The working sample for each MLM model varied from the total number of individuals observed in the data (489,468) according to whether or not they were entered for the subject in question at the particular level. It follows that the largest working samples were for SGI English (378,254) and Maths (284,325) as these subjects are compulsory at this level. The working samples were comprised of those individuals who had taken the specific qualifications *on-time*; that is SGI qualifications in S4, Highers in S5 and Advanced Highers in S6. School level cluster-robust standard errors were employed for all models to account for the possibility that model errors for individuals in the same school might be correlated.

2.5 Results and Analysis

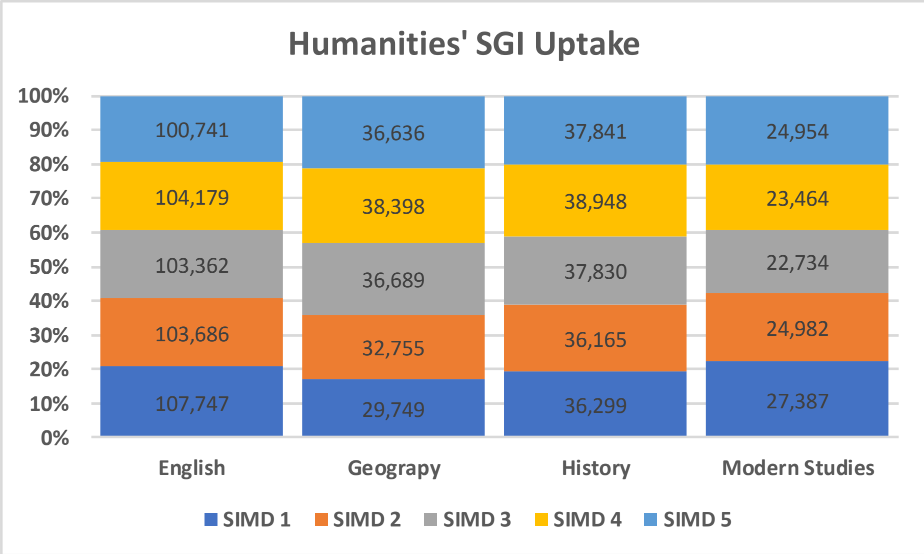
The uptake of Facilitating Subjects by SIMD at SGI, H and AH levels respectively is shown in Figures 2.2, 2.3 and 2.4 below. As previously outlined, the curriculum structure underpinning age-16 qualifications was comprised of a combination of eight subjects: two compulsory subjects (English and Maths), five closed option choices (whereby pupils had to choose a science, humanity, modern language, technology and creative subject) and one completely free option choice that could be used to study, for instance, a second science, humanity or modern language. It should be expected, therefore, that pupils from different SIMD backgrounds would be distributed across subjects fairly evenly at this qualification level. Figure 2.2 shows this to be the case, although some drift is noticeable in the named sciences. For instance, the proportion of pupils from SIMD 1 (the most deprived 20% of households) taking SGI Physics was only half that of those from SIMD 5 (the most affluent 20% of households). General science was available as an alternative choice to a named science and, as shown in Chapter Three, individuals from the two lower SIMDs are overrepresented here.

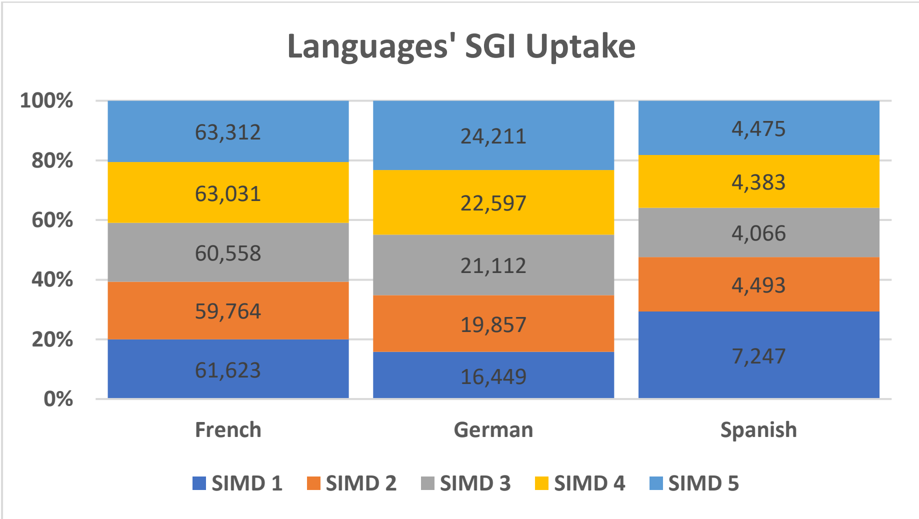
Access to Higher courses was based on SGI attainment (normally Credit Standard Grade 1 or 2 was required for entry) and teacher judgement but was otherwise free choice. It can be seen from Figure 2.3 that individuals from SIMDs 1 and 2, the most deprived 40% of households are underrepresented across all facilitating subjects at Higher, averaging a little over 20% of the uptake across the board. At Advanced Higher, the proportion of SIMD 1 and 2 pupils taking facilitating subjects falls below 20%, with those from SIMD 1 in single figures (Figure 2.4).

The multinomial logit estimation results for attainment in facilitating subjects are presented below in two groups for each level of qualification; Humanities and Languages (English, Geography, History, Modern Studies, French, German, Spanish) and Maths and Science (Biology, Chemistry, Physics). In addition, estimates for the top 50% of achievers at SGI level nationally are provided for each grouping at each qualification level.

Figure 2.2

Facilitating Subject Uptake at Standard Grade/Intermediate 2





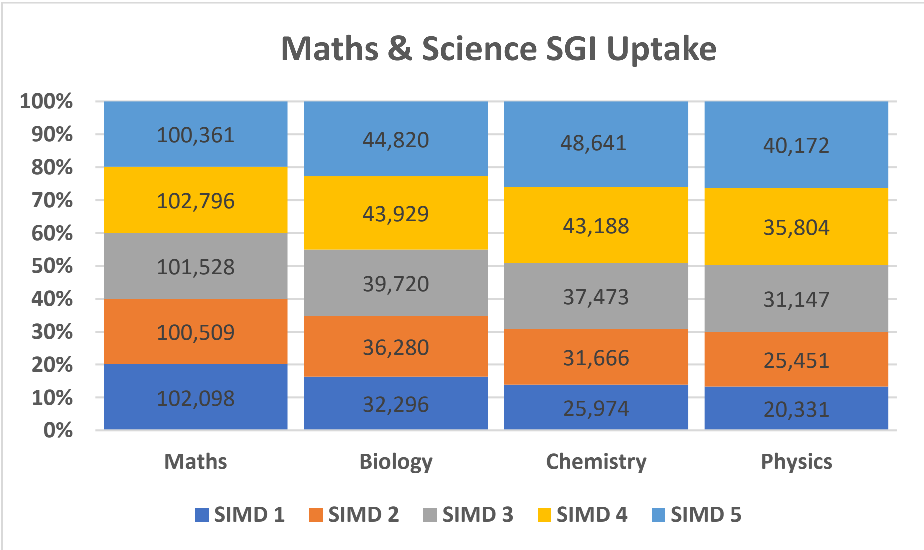


Figure 2.3

Facilitating Subject Uptake at Higher

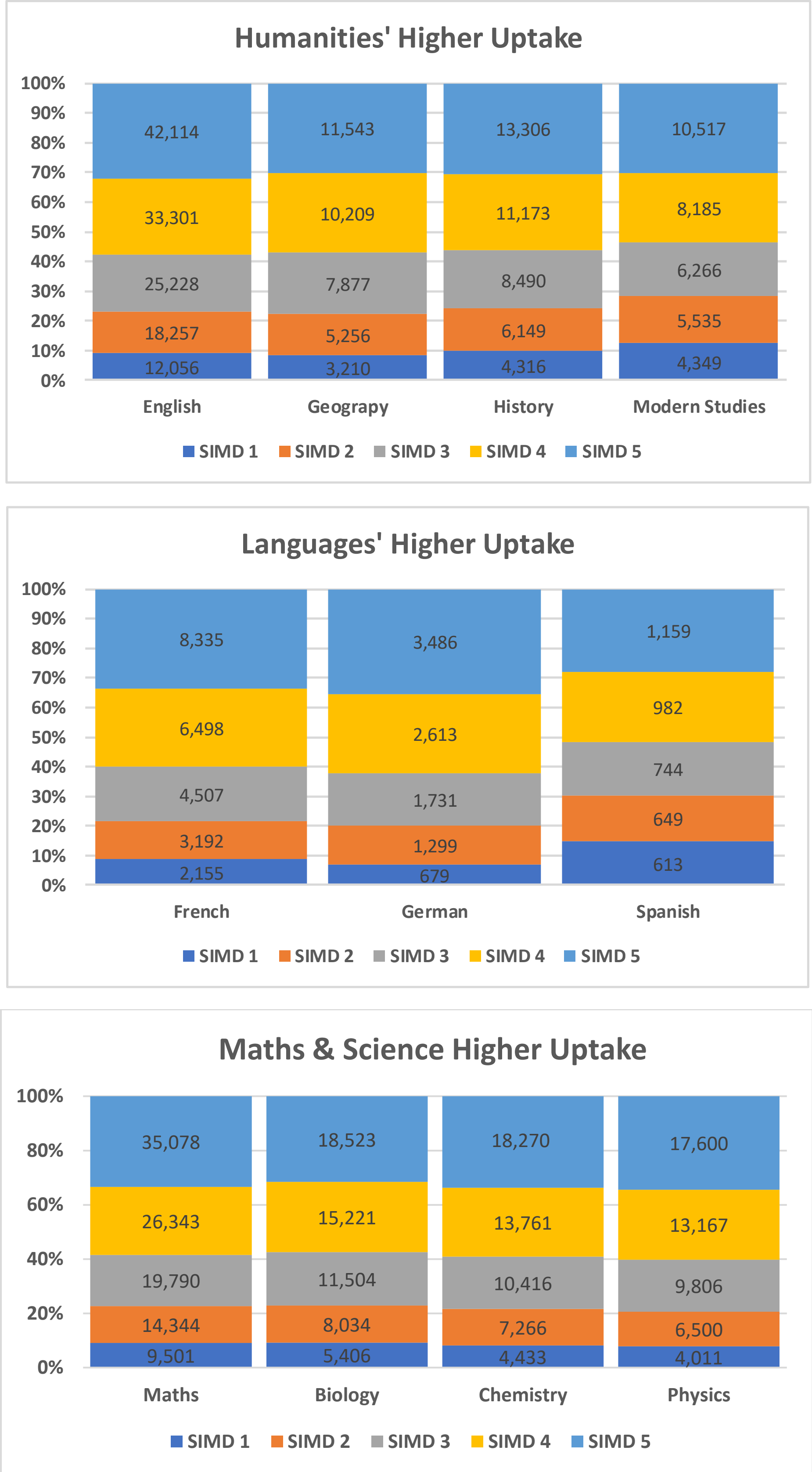
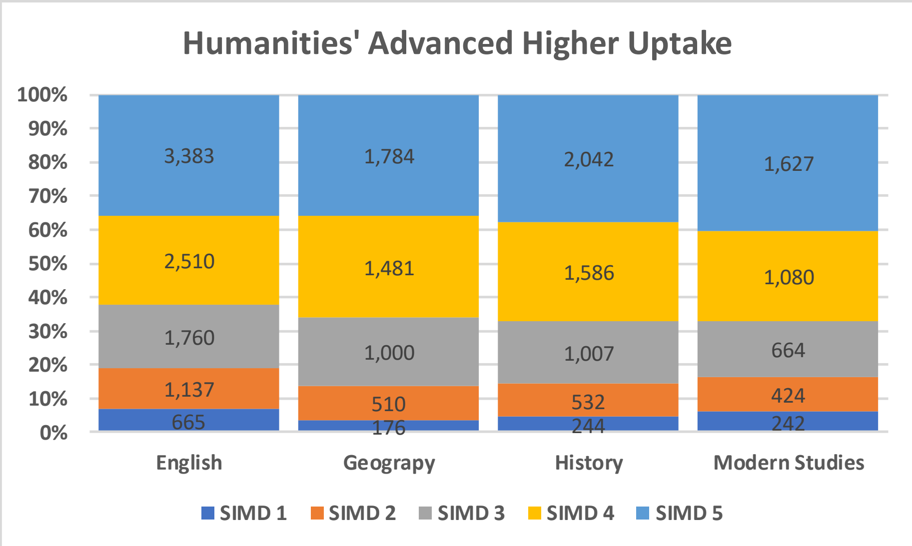
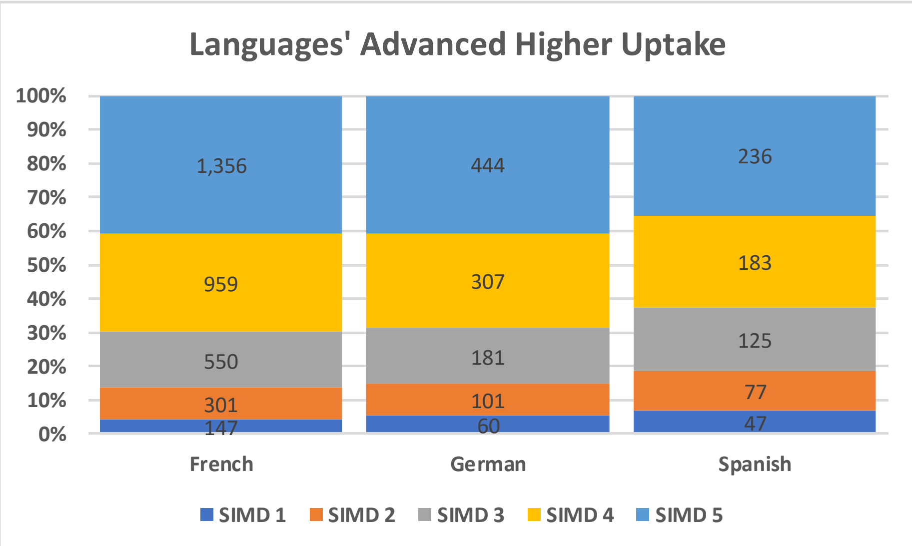
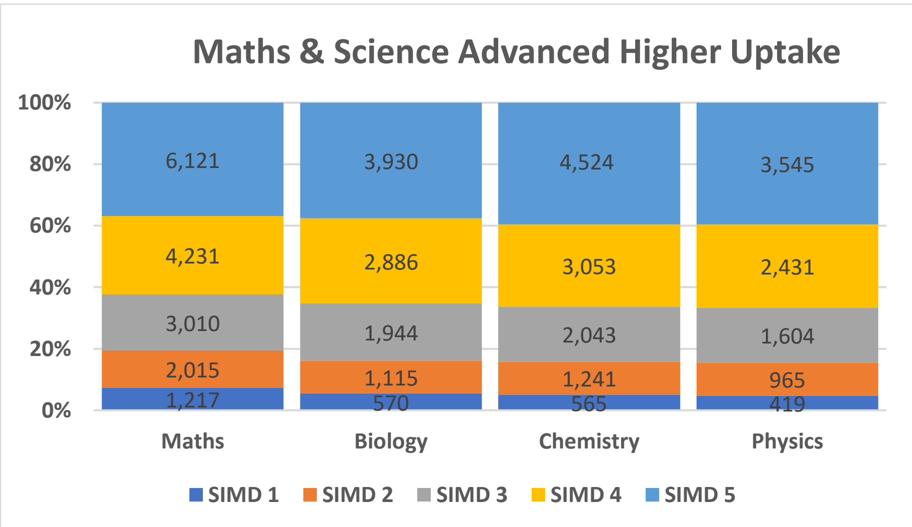


Figure 2.4

Facilitating Subject Uptake at Advanced Higher







2.5.1 Standard Grade / Intermediate 2 Attainment

Tables 2.3 and 2.4 show the Humanities and Languages and Maths and Science SGI results for all students from 2002 to 2009 for the main variables of interest. In Humanities and languages (Table 2.3), in general, males are less likely to pass at all levels rather than fail compared to females. As pass grades rise, they are increasingly less likely to pass as opposed to fail. For example, in History, compared to females, males are 74% as likely to obtain a low pass, 57% as likely to achieve a middle pass and 43% as likely to obtain a high pass. The exceptions to this are English and German at low pass, where males are 36% and 14% respectively more likely to achieve this than females compared to failing and Geography at both low and middle pass, where males are 25% and 11% respectively more likely to obtain these grades. A similar pattern is found for Maths and Science SGI attainment (Table 2.4) with, in general, males increasingly less likely to achieve a pass at the higher grades (as opposed to fail) compared to females. Males are, however, 5% more likely than females to achieve either a low or middle pass in Maths and 17% more likely to obtain a low pass in Physics; although this falls away to being 89% as likely to achieve a middle pass and only 60% as likely to obtain a high pass (the lowest odds of achieving a high pass in a subject in this group). Average marginal effects for both groups of subjects are reported in Tables 2.5 and 2.6. Being male increased the probability of failing by 0.3 of one percent (English) to 3% (Spanish). The probability of obtaining a low pass increased by 1% (Maths/Chemistry) to 15% (French/Spanish) for males compared to females, while the probability of a high pass was reduced by 1% (Maths/Biology) to 11% in Languages. While, in general, males did less well than females at all subjects at SGI level, they did less badly in Maths and Science.

The impact of individuals' socio-economic background, as measured by their SIMD quintile, is stark. Across all subjects, the likelihood of obtaining any passing grade, as opposed to failing, fell as SIMD quintiles moved sequentially from 5 - the least deprived 20% of households, to 1 – the most deprived 20% of households. For the core compulsory subjects of English and Maths, individuals from SIMD 1, compared to those from SIMD 5, were just over 50% as likely to obtain a low pass (as opposed to failing) and only 3% (English) and 6% (Maths) as likely to achieve a high pass. In general, the likelihood of individuals from SIMD 1, compared to those from SIMD 5, obtaining low or high passes in the other subjects ranged from approximately 40% (languages) to 71% (History) for a low pass and 5% (French/German) to 10% (sciences) for a high pass. Average marginal effects (Tables 2.5 and 2.6) indicated that, compared to individuals from SIMD 5, those from lower SIMD households had an increased probability of failing facilitating subjects at SGI level or obtaining a low pass, and, in general a reduced probability of achieving middle or high passes. For individuals from SIMD 1 households, on average, the probability of failure was increased by some 3% (English) to 5% (Geography) while that of obtaining a low pass was increased by 12% (Chemistry/Physics) to 24%-25% (Maths/English). The probability of achieving high passes for SIMD 1 individuals was reduced by 17% (English/Spanish) to 25% (Modern Studies/Physics). In general, the probability of SIMD 1 individuals obtaining middle passes (compared to those from SIMD 5) was reduced except for the sciences where they had an increased probability of obtaining this grade of 1%-8%. This pattern of an increased probability of a middle pass in the sciences but a decreased probability of this grade in all other subjects was repeated for SIMDs 2-4 also.

School size, as measured by SGI cohort size, was seen to have mixed effects according to subject. For English, French and Chemistry at low pass, there was a small but significant, negative impact on the likelihood of obtaining this grade; this fell by 0.1 of one percent for a one-pupil increase in SGI cohort size. There were positive effects, however, of similar magnitude for other subjects; Modern Studies at middle and high grades and Geography, History and Spanish at high grade only. The impact of a school's socio-economic composition, as measured by the percentage of pupils from the most deprived 40% of households (SIMDs 1 and 2), was generally negative but small. For a one percent increase in the proportion of pupils from SIMDs 1 and 2, this ranged from approximately a 1% reduction in the likelihood of achieving a low or middle pass to a 2% fall in the probability of obtaining a high pass.

Table 2.3

Humanities & Languages Standard Grade/Intermediate 2 Attainment 2002-2009 : Population

(omitted category: Outcome 1 Fail)

	English	Geography	History	Modern Studies	French	German	Spanish
Outcome 2: Pass low							
Sex (Male=1)	1.363***	1.252***	0.748***	0.843***	0.948	1.145**	0.908
SIMD 1	0.531***	0.525***	0.706***	0.679***	0.389***	0.434***	0.401***
SIMD 2	0.705***	0.652***	0.790***	0.773**	0.508***	0.527***	0.421***
SIMD 3	0.847*	0.704***	0.880*	0.852*	0.570***	0.629***	0.544***
SIMD 4	1.025	0.888*	0.955	1.009	0.786***	0.806**	0.873
SGI cohort size	0.999*	1.000	1.000	1.001	0.999**	1.000	1.001
% pupils SIMD 1 & 2	0.999	0.992***	0.995***	0.995**	0.996**	0.997*	1.004
Outcome 3: Pass middle							
Sex (Male=1)	0.678***	1.109***	0.570***	0.633***	0.482***	0.595***	0.405***
SIMD 1	0.113***	0.175***	0.222***	0.232***	0.132***	0.150***	0.143***
SIMD 2	0.214***	0.279***	0.339***	0.343***	0.223***	0.236***	0.199***
SIMD 3	0.344***	0.392***	0.464***	0.475***	0.319***	0.313***	0.311***
SIMD 4	0.580***	0.612***	0.641***	0.716***	0.555***	0.526***	0.656*
SGI cohort size	0.999	1.001	1.001	1.002*	0.999	1.001	1.002
% pupils SIMD 1 & 2	0.993***	0.988***	0.990***	0.991***	0.993***	0.992**	0.996
Outcome 4: Pass high							
Sex (Male=1)	0.387***	0.756***	0.435***	0.516***	0.275***	0.320***	0.241***
SIMD 1	0.027***	0.058***	0.080***	0.091***	0.045***	0.053***	0.065***
SIMD 2	0.080***	0.135***	0.161***	0.172***	0.103***	0.109***	0.121***
SIMD 3	0.175***	0.233***	0.267***	0.290***	0.190***	0.174***	0.208***
SIMD 4	0.400***	0.464***	0.456***	0.550***	0.405***	0.400***	0.541**
SGI cohort size	1.000	1.002*	1.002**	1.003**	1.001	1.001	1.005*
% pupils SIMD 1 & 2	0.986***	0.982***	0.982***	0.984***	0.986***	0.985***	0.983**
Observations	378254	123097	135516	90679	218395	72203	16950
Log pseudolikelihood	-324494.55	-144438.87	-158994.53	-105814.11	-243474.88	-79493.806	-18513.096
Wald chi2 (57)	13857.90***	5393.45***	5648.37***	4656.70***	9051.13***	4440.30***	4916.72***
Pseudo R^2	0.072	0.055	0.058	0.054	0.064	0.060	0.082

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (Clusters by subject: English 355, Geography 351, History 350, Modern Studies 296, French 351, German 275, Spanish 178).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.4

Maths & Science Standard Grade/Intermediate 2 Attainment 2002-2009 : Population

(omitted category: Outcome 1 Fail)

	Maths	**Biology**	**Chemistry**	**Physics**
Outcome 2: Pass low				
Sex (Male=1)	1.052***	0.967	0.883**	1.166***
SIMD 1	0.524***	0.569***	0.675***	0.631***
SIMD 2	0.622***	0.640***	0.780*	0.755***
SIMD 3	0.714***	0.774***	0.877	0.771***
SIMD 4	0.843***	0.856*	1.119	0.899
SGI cohort size	1.000	1.000	0.999*	1.000
% pupils SIMD 1 & 2	0.996***	1.001	1.004	0.999
Outcome 3: Pass middle				
Sex (Male=1)	1.052**	0.855***	0.846***	0.891**
SIMD 1	0.170***	0.265***	0.276***	0.269***
SIMD 2	0.272***	0.352***	0.393***	0.391***
SIMD 3	0.387***	0.505***	0.546***	0.481***
SIMD 4	0.590***	0.678***	0.821	0.681***
SGI cohort size	1.000	1.001	0.999	1.000
% pupils SIMD 1 & 2	0.993***	0.997	1.001	0.996*
Outcome 4: Pass high				
Sex (Male=1)	0.949**	0.834***	0.785***	0.598***
SIMD 1	0.063***	0.099***	0.099***	0.098***
SIMD 2	0.133***	0.171***	0.188***	0.185***
SIMD 3	0.234***	0.325***	0.331***	0.286***
SIMD 4	0.442***	0.509***	0.609***	0.515***
SGI cohort size	1.000	1.001	1.000	1.000
% pupils SIMD 1 & 2	0.989***	0.986***	0.991***	0.986***
Observations	361671	142151	133749	108595
Log pseudolikelihood	-434107.59	-156790.78	-137889.13	-117522.18
Wald chi2(57)	6498.47***	3173.36***	3662.68***	3690.74***
Pseudo R^2	0.046	0.039	0.040	0.042

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant. SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (Clusters by subject: Maths 354, Biology 352, Chemistry 352, Physics 352).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.000$

Table 2.5

Humanities & Languages Attainment at Standard Grade / Intermediate 2 : Average Marginal Effects for the Population

Variable	Grade	English	Geography	History	Modern Studies	French	German	Spanish
Sex (M=1)	Fail	0.003***	-0.006**	0.034***	0.019***	0.026***	0.019***	0.034***
	Low pass	0.116***	0.040***	0.049***	0.058***	0.145***	0.134***	0.145***
	Medium pass	-0.052***	0.030***	-0.015***	-0.018***	-0.062***	-0.039***	-0.066***
	High pass	-0.067***	-0.064***	-0.067***	-0.060***	-0.109***	-0.114***	-0.113***
SIMD 1	Fail	0.027***	0.121***	0.085***	0.062***	0.087***	0.084***	0.084***
	Low pass	0.254***	0.192***	0.220***	0.234***	0.222***	0.201***	0.175***
	Medium pass	-0.113***	-0.071***	-0.053***	-0.042***	-0.104***	-0.087***	-0.091***
	High pass	-0.168***	-0.242***	-0.252***	-0.255***	-0.205***	-0.197***	-0.168***
SIMD 2	Fail	0.019***	0.086***	0.061***	0.045***	0.064***	0.064***	0.071***
	Low pass	0.194***	0.144***	0.161***	0.177***	0.167***	0.151***	0.120***
	Medium pass	-0.095***	-0.065***	-0.038***	-0.033***	-0.082***	-0.067***	-0.078***
	High pass	-0.118***	-0.165***	-0.184***	-0.189***	-0.149***	-0.147***	-0.113***
SIMD 3	Fail	0.013***	0.064***	0.043***	0.031***	0.049***	0.050***	0.051***
	Low pass	0.146***	0.099***	0.123***	0.127***	0.115***	0.129***	0.092***
	Medium pass	-0.077***	-0.045***	-0.028***	-0.024***	-0.061***	-0.064***	-0.054***
	High pass	-0.082***	-0.118***	-0.137***	-0.134***	-0.103***	-0.116***	-0.088***
SIMD 4	Fail	0.006***	0.032***	0.024***	0.013***	0.024***	0.027***	0.017
	Low pass	0.092***	0.063***	0.077***	0.075***	0.071***	0.077***	0.050**
	Medium pass	-0.051***	-0.030***	-0.017**	-0.014*	-0.034***	-0.045***	-0.025
	High pass	-0.046***	-0.065***	-0.084***	-0.073***	-0.061***	-0.058***	-0.042**

Notes:
Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one Unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6

Maths & Science Attainment at Standard Grade / Intermediate 2 : Average Marginal Effects for the Population

Variable	Grade	Maths	Biology	Chemistry	Physics
Sex	Fail	-0.003*	0.007***	0.006***	0.008***
	Low pass	0.008***	0.017***	0.006**	0.043***
	Medium pass	0.009***	-0.013***	0.006*	0.039***
	High pass	-0.014***	-0.010**	-0.018***	-0.090***
SIMD 1	Fail	0.106***	0.067***	0.046***	0.070***
	Low pass	0.243***	0.132***	0.123***	0.119***
	Medium pass	-0.122***	0.014*	0.075***	0.059***
	High pass	-0.227***	-0.212***	-0.243***	-0.248***
SIMD 2	Fail	0.078***	0.052***	0.033***	0.050***
	Low pass	0.178***	0.101***	0.093***	0.092***
	Medium pass	-0.091***	0.004	0.051***	0.043***
	High pass	-0.165***	-0.157***	-0.177***	-0.184***
SIMD 3	Fail	0.056***	0.033***	0.021***	0.038***
	Low pass	0.131***	0.070***	0.064***	0.064***
	Medium pass	-0.069***	-0.005	0.034***	0.026***
	High pass	-0.118***	-0.099***	-0.120***	-0.129***
SIMD 4	Fail	0.031***	0.019***	0.007*	0.020***
	Low pass	0.077***	0.040***	0.042***	0.037***
	Medium pass	-0.040***	0.003	0.022***	0.013*
	High pass	-0.068***	-0.062***	-0.071***	-0.070***

Notes:
Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The SGI pattern of gender attainment in humanities and languages for the whole sample is replicated amongst the top 50% of achievers nationally with boys being more likely to obtain a low pass in English and a low or middle pass in Geography but otherwise performing less well than girls (Table 2.7). Comparison of the average marginal effects for both the population (Table 2.5) and the top performers (Table 2.9) confirmed this and demonstrated that, in general, compared to females, males have an increased probability of failing Humanities and Languages subjects at SGI level or obtaining a low pass and a reduced probability of some 6% - 12% of achieving a high pass. The multinomial logit regression results indicated that top performing boy were more likely to obtain Maths passes at all levels (Table 2.8), as opposed to a low or middle pass in Maths among the whole sample. On average, however, boys had a reduced probability of achieving a high pass in Maths and/or Science (Tables 2.6 and 2.10) and, again, increased probabilities of fail or low pass compared to girls.

At the low pass level in both subject groups, the effect of socio-economic background for the top 50% of achievers was far less than in the whole sample and appeared to be subject specific. For Geography, French and German, individuals from lower SIMDs were less likely to achieve a low pass (compared to failing) than those from SIMD 5, the least deprived 20% of households. In the core subject of English, those from SIMDs 1 and 2, the most deprived 40% of households, were as likely to obtain a low pass compared to those from SIMD 5, but only 10% and 17% respectively as likely to achieve a high pass compared to the latter. Socio-economic effects became prevalent across all humanities and languages subjects at middle and high pass for the top 50% of achievers although the magnitudes of these were less than for the whole sample. This is confirmed by comparison of the average marginal effects for Humanities and Languages for the population and the top 50% (Tables 2.5 and 2.9). Individuals from SIMD 1 and 2 households had an increased probability of failing or obtaining a low pass (compared to those from SIMD 5) but the impact was reduced for the top 50%. Compared to the population as a whole, the top 50% had an increased probability of obtaining a middle pass in the humanities but not in languages, and a similar reduction in the probability of achieving a high pass across the board. These effects displayed the established pattern of impact whereby, the lower an individual's SIMD, the greater the reduction in their probability of obtaining one of the higher grades.

For the top 50% of achievers in Maths and science, those from SIMD 1 were less likely than those from SIMD 5 to achieve a low pass in Maths, Biology or Physics (compared to failing) but no less likely to obtain a Chemistry low pass (Table 2.8). At middle and high pass, individuals from all lower SIMDs were significantly less likely to achieve these grades than those from SIMD 5, with odds reducing sequentially as SIMD fell, although, as with humanities and languages subjects, the magnitude of the effects was less than for the whole sample. A top 50% individual from SIMD 4 was 58% as likely as someone from SIMD 5 to obtain a Maths' high pass whilst a top 50% individual from SIMD 1 was only 14% as likely to achieve this. Compared to individuals from SIMD 5 households, on average, those from lower SIMDs had a reduced probability of achieving a high pass in Maths and Science subjects (Table 2.10); the lower an individual's SIMD, the greater the reduction in that probability. The average marginal effects for the top 50% were generally lower than those for the population at the fail, low and middle pass outcomes but of similar magnitude at high pass with a 21%-25% reduction in the probability of achieving such in Maths and Science for an SIMD 1 individual compared to an SIMD 5 individual.

School size and socio-economic composition effects were less prevalent but generally of similar magnitude for the top 50% of achievers with the notable exception of English at low and middle pass (Tables 2.7 and 2.8). Here, for a 1% increase in the percentage of pupils from SIMDs 1 and 2 (the most deprived 40%), the likelihood of obtaining a low or middle pass in English rose by 1%. Small positive effects for schools' socio-economic composition were also seen at low pass for Biology and Chemistry. For a 1% increase in the percentage of pupils from SIMDs 1 and 2, the likelihood of obtaining such increased by approximately 0.4% to 0.8%.

Table 2.7

Humanities & Languages Standard Grade/Intermediate 2 Attainment 2002-2009 : Individuals with Above Median SGI Points

(omitted category: Outcome 1 Fail)

	English	Geography	History	Modern Studies	French	German	Spanish
Outcome 2: Pass low							
Sex (Male=1)	1.496**	1.411***	0.800***	0.828**	0.787***	1.074	0.863
SIMD 1	1.224	0.643***	1.099	0.996	0.635***	0.587***	0.578
SIMD 2	1.109	0.785**	1.075	1.024	0.667***	0.644***	0.598
SIMD 3	1.595*	0.766**	1.068	0.990	0.689***	0.778*	0.679
SIMD 4	1.504	0.925	1.091	1.023	0.827*	0.829	0.905
SGI cohort size	1.001	1.000	0.999	1.000	0.999	1.000	1.001
% pupils SIMD 1 & 2	1.016***	0.992***	0.997	0.996	1.001	1.002	1.005
Outcome 3: Pass middle							
Sex (Male=1)	0.606***	1.372***	0.643***	0.631***	0.359***	0.489***	0.337***
SIMD 1	0.381***	0.268***	0.430***	0.428***	0.265***	0.244***	0.224***
SIMD 2	0.437***	0.383***	0.545***	0.542***	0.336***	0.327***	0.310***
SIMD 3	0.775	0.478***	0.642***	0.620***	0.419***	0.422***	0.398***
SIMD 4	0.955	0.694***	0.793**	0.779	0.610***	0.571***	0.696
SGI cohort size	1.001	1.001	1.000	1.001	0.999	1.000	1.002
% pupils SIMD 1 & 2	1.012**	0.987***	0.993***	0.993*	0.998	0.999	0.998
Outcome 4: Pass high							
Sex (Male=1)	0.340***	0.945	0.492***	0.516***	0.204***	0.260***	0.198***
SIMD 1	0.098***	0.090***	0.160***	0.172***	0.092***	0.088***	0.101***
SIMD 2	0.170***	0.188***	0.264***	0.276***	0.157***	0.153***	0.188***
SIMD 3	0.406***	0.286***	0.372***	0.381***	0.250***	0.236***	0.261***
SIMD 4	0.661*	0.530***	0.571***	0.602***	0.447***	0.437***	0.561
SGI cohort size	1.002	1.002*	1.001	1.002	1.001	1.001	1.005*
% pupils SIMD 1 & 2	1.005	0.981***	0.984***	0.985***	0.991**	0.991**	0.986*
Observations	288853	105080	113165	76520	189053	63272	14980
Log pseudolikelihood	-198185.51	-110040.61	-119204.1	-81122.422	-192661.34	-62935.409	-14894.79
Wald chi2(57, Spanish 56)	8810.62***	3477.69***	3590.75***	3224.78***	8092.22***	3848.07***	-
Pseudo R^2	0.059	0.045	0.044	0.042	0.056	0.054	0.079

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (Clusters by subject: English 353, Geography 351, History 350, Modern Studies 289, French 351, German 269, Spanish 158).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.8

Maths & Science Standard Grade/Intermediate 2 Attainment 2002-2009 : Individuals with Above Median SGI Points

(omitted category: Outcome 1 Fail)

	Maths	Biology	Chemistry	Physics
Outcome 2: Pass low				
Sex (Male=1)	1.429***	1.064	1.024	1.327***
SIMD 1	0.864*	0.744***	0.928	0.762***
SIMD 2	0.907	0.803**	0.951	0.866
SIMD 3	0.952	0.959	1.005	0.835*
SIMD 4	1.037	0.924	1.262	0.920
SGI cohort size	1.000	1.000	0.999	1.000
% pupils SIMD 1 & 2	0.997*	1.004*	1.008*	1.000
Outcome 3: Pass middle				
Sex (Male=1)	1.644***	0.974	1.018	1.077
SIMD 1	0.360***	0.374***	0.409***	0.344***
SIMD 2	0.473***	0.461***	0.500***	0.466***
SIMD 3	0.589***	0.643***	0.639***	0.535***
SIMD 4	0.773***	0.741**	0.944	0.709***
SGI cohort size	1.000	1.001	1.000	1.000
% pupils SIMD 1 & 2	0.993***	1.001	1.006	0.997
Outcome 4: Pass high				
Sex (Male=1)	1.488***	0.955	0.949	0.725***
SIMD 1	0.135***	0.140***	0.148***	0.126***
SIMD 2	0.231***	0.226***	0.241***	0.221***
SIMD 3	0.356***	0.415***	0.388***	0.320***
SIMD 4	0.580***	0.556***	0.702**	0.536***
SGI cohort size	1.001	1.002**	1.001	1.001
% pupils SIMD 1 & 2	0.988***	0.990***	0.997	0.987***
Observations	284325	133430	128679	104383
Log pseudolikelihood	-311478.36	-136924.7	-125418.54	-107680.11
Wald chi2(57)	4119.67***	2390.12***	2850.61***	3270.52***
Pseudo R^2	0.036	0.035	0.036	0.042

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant. SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households. Other independent variables included: Urban/Rural location and Year fixed effects. Cluster-robust standard errors at the school level (Clusters by subject: Maths 353, Biology 351, Chemistry/Physics 352). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.000$

Table 2.9

Humanities & Languages Attainment at Standard Grade / Intermediate 2 :

Average Marginal Effects for Individuals with Above Median SGI points

Variable	Grade	English	Geography	History	Modern Studies	French	German	Spanish
Sex (M=1)	Fail	0.000**	-0.006***	0.010***	0.006***	0.011***	0.009***	0.015***
	Low pass	0.084***	0.021***	0.041***	0.052***	0.164***	0.148***	0.163***
	Medium pass	-0.007**	0.050***	0.014***	0.000	-0.054***	-0.030***	-0.057***
	High pass	-0.077***	-0.066***	-0.064***	-0.058***	-0.121***	-0.126***	-0.121***
SIMD 1	Fail	0.001**	0.033***	0.019***	0.011***	0.016***	0.019***	0.022***
	Low pass	0.114***	0.167***	0.174***	0.182***	0.200***	0.174***	0.172***
	Medium pass	0.056***	0.036***	0.047***	0.045***	-0.016*	-0.003	-0.030
	High pass	-0.171***	-0.235***	-0.239***	-0.238***	-0.200***	-0.190***	-0.164***
SIMD 2	Fail	0.001***	0.023***	0.013***	0.008***	0.013***	0.015***	0.017***
	Low pass	0.090***	0.130***	0.127***	0.136***	0.154***	0.133***	0.117***
	Medium pass	0.030***	0.007	0.035***	0.033***	-0.020**	-0.005	-0.028
	High pass	-0.120***	-0.160***	-0.175***	-0.177***	-0.147***	-0.143***	-0.106***
SIMD 3	Fail	0.000	0.018***	0.010***	0.006***	0.011***	0.011***	0.014**
	Low pass	0.069***	0.087***	0.095***	0.099***	0.109***	0.117***	0.096***
	Medium pass	0.014**	0.009	0.027***	0.023**	-0.019**	-0.016	-0.021
	High pass	-0.083***	-0.113***	-0.132***	-0.128***	-0.101***	-0.113***	-0.089***
SIMD 4	Fail	0.000	0.009***	0.005**	0.003	0.006***	0.007***	0.005
	Low pass	0.043***	0.052***	0.059***	0.057***	0.067***	0.068***	0.048***
	Medium pass	0.005	0.001	0.015*	0.009	-0.012*	-0.020*	-0.009
	High pass	-0.048***	-0.061***	-0.080***	-0.069***	-0.061***	-0.056***	-0.045**

Notes:
Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.10

Maths & Science Attainment at Standard Grade / Intermediate 2 :

Average Marginal Effects for Individuals with Above Median SGI points

Variable	Grade	Maths	Biology	Chemistry	Physics
Sex	Fail	-0.008***	0.000	0.000	0.000
	Low pass	-0.019***	0.012***	0.003	0.036***
	Medium pass	0.033***	-0.006	0.012***	0.051***
	High pass	-0.006**	-0.007*	-0.014***	-0.087***
SIMD 1	Fail	0.018***	0.024***	0.018***	0.038***
	Low pass	0.230***	0.125***	0.112***	0.114***
	Medium pass	-0.030***	0.059***	0.108***	0.090***
	High pass	-0.217***	-0.208***	-0.237***	-0.242***
SIMD 2	Fail	0.013***	0.019***	0.014***	0.027***
	Low pass	0.170***	0.098***	0.085***	0.088***
	Medium pass	-0.023***	0.037***	0.073***	0.066***
	High pass	-0.160***	-0.153***	-0.172***	-0.181***
SIMD 3	Fail	0.009***	0.010***	0.009***	0.022***
	Low pass	0.124***	0.068***	0.060***	0.062***
	Medium pass	-0.020***	0.018**	0.049***	0.042***
	High pass	-0.113***	-0.096***	-0.117***	-0.126***
SIMD 4	Fail	0.004**	0.007***	0.002	0.012***
	Low pass	0.075***	0.039***	0.038***	0.036***
	Medium pass	-0.014***	0.015**	0.030***	0.022**
	High pass	-0.066***	-0.061***	-0.070***	-0.069***

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.5.2 Highers' Attainment

Results for Higher attainment in Humanities and Languages and Maths and Science for all students from 2002 to 2009 for the main variables of interest are shown in Tables 2.11 and 2.12. The effect of gender was seen to lessen for humanities and languages' attainment at Higher level (Table 2.11). For lower passes, the only difference between the sexes was in English, where males were 88% as likely as females to achieve a low pass at Higher compared to failing. In fact, Higher English was the only subject where there were gender differences at all pass grades, with males being less likely to achieve any given pass grade than females. Whilst males were less likely than females to achieve middle or high passes in Geography and German, on average there was no difference at middle pass (Table 2.13). Apart from these, there were no other gender effects in the humanities and languages subjects at H level. The average marginal effect of being male was to reduce the probability of achieving a high pass by 2% (English) to 4% (Geography/German).

For Maths and science at Higher, there were significant gender effects across all subjects, at all pass grades, uniformly in favour of males (Table 2.12). Males were found to be approximately 70% more likely than females to achieve high passes in Maths and Biology (as opposed to failing) and more than twice as likely to obtain high passes in Chemistry and Physics. On average, this translated into a 4% (Biology) to 8% (Chemistry/Physics) increase in the probability of achieving a high pass in Maths or Science (Table 2.14). Compared to females, males had a 3% - 6% reduced probability of failing Maths or Science subjects at Higher as opposed to a 1% increased probability of failure in such at SGI level. This would seem to be in keeping with the findings of Machin and Pekkarinen (2008) whose PISA analysis would suggest that male attainment is bimodal, with higher variability among boys displayed in PISA reading and Maths' tests. Previous attainment as measured by individuals' total SGI UCAS points was significant in all subjects at all pass levels, with the impact ranging from a 2% increase in the likelihood of obtaining a low pass in any subject to an 8% increase in achieving a high pass in Biology/Chemistry for an extra SGI point.

Table 2.11

Humanities & Languages Higher Attainment 2002-2009 : Population

(omitted category: Outcome 1 Fail)

	English	Geography	History	Modern Studies	French	German	Spanish
Outcome 2: Pass low							
Sex (Male=1)	0.878***	0.936	1.058	1.039	0.955	0.897	1.040
SGI UCAS points	1.018***	1.017***	1.015***	1.017***	1.017***	1.017***	1.016***
SIMD 1	0.850***	0.945	0.860*	0.991	0.905	1.154	0.786
SIMD 2	0.919**	0.954	0.892*	1.014	0.902	1.217	1.075
SIMD 3	0.920**	1.016	0.872*	0.985	0.932	0.983	1.115
SIMD 4	0.958	1.052	0.945	1.045	0.884	0.812	1.041
Higher cohort size	1.000	1.001	1.000	1.001	1.000	1.002	0.997
% pupils SIMD 1 & 2	0.996***	0.995***	0.994***	0.994***	0.996	0.995	0.989*
Outcome 3: Pass middle							
Sex (Male=1)	0.774***	0.842***	1.078	1.006	0.895	0.736***	0.993
SGI UCAS points	1.035***	1.034***	1.033***	1.034***	1.035***	1.037***	1.026***
SIMD 1	0.809***	0.821*	0.698***	0.852*	0.884	1.624**	0.789
SIMD 2	0.909*	0.886	0.768***	0.839**	0.943	1.171	1.656*
SIMD 3	0.915*	0.982	0.797***	0.926	0.975	0.978	0.975
SIMD 4	0.941	0.992	0.881	0.993	0.953	0.920	1.095
Higher cohort size	1.000	1.003	1.000	1.001	0.998	1.001	0.999
% pupils SIMD 1 & 2	0.995***	0.995*	0.990***	0.991***	0.994	0.987***	0.996
Outcome 4: Pass high							
Sex (Male=1)	0.703***	0.683***	1.062	0.960	0.908	0.653***	1.089
SGI UCAS points	1.057***	1.061***	1.057***	1.056***	1.062***	1.064***	1.056***
SIMD 1	0.712***	0.648***	0.571***	0.739**	0.720*	1.407	0.905
SIMD 2	0.837**	0.826*	0.691***	0.769**	0.810	1.127	1.323
SIMD 3	0.853**	1.069	0.700***	0.830*	0.917	0.845	1.067
SIMD 4	0.900*	1.105	0.813*	0.933	0.865	0.705**	1.011
Higher cohort size	1.001	1.004	1.000	1.001	0.999	1.002	0.999
% pupils SIMD 1 & 2	0.991***	0.993*	0.985***	0.989***	0.992	0.986**	0.992
Observations	109015	28991	33967	26636	19571	7779	3013
Log pseudolikelihood	-128720.01	-32866.246	-38904.38	-31069.139	-20863.906	-8527.3578	-3097.6502
Wald chi2(69)	8658.09***	4703.55***	5221.90***	4808.02***	1945.21***	1617.71***	7646.50***
Pseudo R^2	0.136	0.181	0.159	0.156	0.184	0.191	0.179

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects. Cluster-robust standard errors at the school level (Clusters by subject: English 341, Geography 337, History 335, Modern Studies 302, French 336, German 256, Spanish 155). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.12

Maths & Science Higher Attainment 2002-2009 : Population

(omitted category: Outcome 1 Fail)

	Maths	Biology	Chemistry	Physics
Outcome 2: Pass low				
SEX (MALE=1)	1.048*	1.179***	1.250***	1.138***
SGI UCAS points	1.018***	1.020***	1.023***	1.019***
SIMD 1	0.850***	1.009	0.889	0.908
SIMD 2	0.931	1.016	0.900	0.939
SIMD 3	0.893***	0.998	1.000	0.952
SIMD 4	0.976	0.982	1.023	0.935
Higher cohort size	1.000	0.999	0.999	1.000
% pupils SIMD 1 & 2	0.998	0.994***	0.995**	0.995**
Outcome 3: Pass middle				
SEX (MALE=1)	1.272***	1.350***	1.687***	1.540***
SGI UCAS points	1.035***	1.044***	1.047***	1.040***
SIMD 1	0.808***	0.973	0.828**	0.799**
SIMD 2	0.884**	0.941	0.849**	0.839**
SIMD 3	0.867**	1.009	0.895	0.861*
SIMD 4	0.949	0.955	0.969	0.906
Higher cohort size	1.000	0.999	0.999	1.000
% pupils SIMD 1 & 2	0.998	0.992***	0.992***	0.994**
Outcome 4: Pass high				
SEX (MALE=1)	1.690***	1.763***	2.455***	2.104***
SGI UCAS points	1.061***	1.077***	1.077***	1.070***
SIMD 1	0.645***	0.754**	0.616***	0.610***
SIMD 2	0.755***	0.850*	0.714***	0.798**
SIMD 3	0.780***	0.867	0.772***	0.790**
SIMD 4	0.886*	0.869	0.901	0.880
Higher cohort size	1.000	0.998	1.000	1.001
% pupils SIMD 1 & 2	0.995	0.988***	0.989***	0.986***
Observations	80817	45556	42520	39480
Log pseudolikelihood	-96576.154	-50546.733	-48048.978	-44919.408
Wald chi2(69)	4351.39***	5153.08***	3777.63***	4348.56***
Pseudo R^2	0.136	0.198	0.184	0.173

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant. SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households. Other independent variables included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects. Cluster-robust standard errors at the school level (Clusters by subject: Maths 347, Biology 341, Chemistry 342, Physics 342).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.13

Humanities & Languages Attainment at Higher :

Population Average Marginal Effects Average Marginal Effects of Gender & Socio-economic Background

Variable	Grade	English	Geography	History	Modern Studies	French	German	Spanish
Sex (M=1)	Fail	0.031***	0.021***	-0.008	-0.002	0.008	0.029**	-0.004
	Low pass	0.007**	0.015**	0.002	0.009	0.003	0.019	0.000
	Medium pass	-0.016***	0.001	0.006	0.001	-0.008	-0.010	-0.010
	High pass	-0.022***	-0.038***	0.000	-0.008	-0.004	-0.037***	0.013
SGI UCAS points	Fail	-0.004***	-0.004***	-0.003***	-0.004***	-0.003***	-0.004***	-0.003***
	Low pass	-0.001***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***
	Medium pass	0.002***	0.001***	0.001***	0.001***	-0.001***	0.000	-0.001***
	High pass	0.004***	0.005***	0.004***	0.004***	0.006***	0.006***	0.006***
SIMD 1	Fail	0.031***	0.023*	0.033***	0.014	0.017	-0.035	0.018
	Low pass	-0.003	0.021*	0.023*	0.023*	0.009	-0.017	-0.016
	Medium pass	-0.007	0.000	-0.020	-0.008	0.014	0.046*	-0.019
	High pass	-0.021***	-0.044**	-0.037***	-0.029*	-0.039*	0.007	0.017
SIMD 2	Fail	0.016**	0.013	0.024***	0.012	0.012	-0.020	-0.026
	Low pass	-0.003	0.007	0.015	0.026**	-0.001	0.017	-0.026
	Medium pass	-0.001	-0.005	-0.017	-0.015	0.014	0.007	0.051*
	High pass	-0.012*	-0.016	-0.022*	-0.024**	-0.025*	-0.004	0.001
SIMD 3	Fail	0.015**	-0.002	0.024***	0.009	0.006	0.006	-0.005
	Low pass	-0.004	0.001	0.007	0.011	-0.004	0.007	0.011
	Medium pass	-0.001	-0.009	-0.009	0.000	0.007	0.011	-0.014
	High pass	-0.010*	0.010	-0.022**	-0.019*	-0.009	-0.024	0.007
SIMD 4	Fail	0.009*	-0.006	0.012	-0.002	0.011	0.024	-0.004
	Low pass	0.000	0.005	0.008	0.012	-0.008	-0.011	0.001
	Medium pass	-0.002	-0.011	-0.006	0.001	0.011	0.026*	0.012
	High pass	-0.007	0.012	-0.014*	-0.011	-0.014	-0.039**	-0.008

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.14

Maths & Science Attainment at Higher :

Population Average Marginal Effects of Gender & Socio-economic Background

Variable	Grade	Maths	Biology	Chemistry	Physics
Sex	Fail	-0.033***	-0.037***	-0.055***	-0.050***
	Low pass	-0.030***	-0.010*	-0.034***	-0.035***
	Medium pass	0.003	0.004	0.009*	0.009
	High pass	0.061***	0.044***	0.080***	0.075***
SGI points	Fail	-0.005***	-0.005***	-0.005***	-0.005***
	Low pass	-0.002***	-0.002***	-0.002***	-0.002***
	Medium pass	0.001***	0.001***	0.001***	0.001***
	High pass	0.006***	0.006***	0.006***	0.006***
SIMD 1	Fail	0.037***	0.006	0.026**	0.031***
	Low pass	0.005	0.014	0.013	0.018*
	Medium pass	0.002	0.014	0.011	0.005
	High pass	-0.044***	-0.033***	-0.051***	-0.055***
SIMD 2	Fail	0.021***	0.005	0.021**	0.019*
	Low pass	0.008	0.013	0.007	0.009
	Medium pass	0.002	-0.001	0.004	-0.010
	High pass	-0.031***	-0.017*	-0.032***	-0.018*
SIMD 3	Fail	0.024***	0.002	0.010	0.017*
	Low pass	0.000	0.004	0.019**	0.010
	Medium pass	-0.001	0.011	0.001	-0.005
	High pass	-0.022***	-0.018**	-0.029***	-0.022**
SIMD 4	Fail	0.009	0.006	0.002	0.012
	Low pass	0.005	0.005	0.011	0.000
	Medium pass	0.001	0.003	0.001	-0.004
	High pass	-0.014*	-0.014*	-0.014*	-0.009

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Socio-economic effects at the low pass grade in humanities and languages (Table 2.11) were much less in evidence than at SGI level. As pass grades rose, they became more prevalent, becoming significant for those from SIMDs 1 and 2 in Geography and Modern Studies in addition to English and History. The general pattern of impact, whereby the lower an individual's SIMD, the lower their likelihood of obtaining a given pass in a subject, was evident. For individuals from SIMD 1 backgrounds, the most deprived 20% of households, the likelihood of their obtaining a high pass in a humanities or language subject, compared to those from SIMD 5, ranged from 57% in History to 74% in Modern Studies. Average marginal effects indicated that being from an SIMD 1 household increased an individual's probability of a low pass in Geography, History and Modern Studies by 2% and reduced their probability of a high pass by 2% - 4% for Humanities subjects (Table 2.13). SIMD effects were virtually absent for languages.

Similarly, for Maths and science (Table 2.12) at low pass, socio-economic background effects were much less in evidence (Maths only with no consistent pattern of impact). Again, as pass grades rose, SIMD effects became more prevalent; in Maths, Chemistry and Physics at middle pass and in all four subjects at high pass, with the likelihood of achieving either grade generally falling with an individual's SIMD. On average, compared to individuals from SIMD 5 households, those from SIMD 1 had a reduced probability of achieving a high pass in Maths and/or a Science at Higher of 3% to 6% (Table 2.14).

The size of a school's Highers' cohort was insignificant for all subjects at all pass levels. Schools' socio-economic composition generally exerted a small negative impact across subjects at most pass grades with the exception of Maths and French (where there was no effect at any level). For a 1% increase in a school's percentage of pupils from SIMDs 1 and 2 (the most deprived 40% of households) the likelihood of obtaining a given pass in a facilitating subject at Higher was reduced by between 0.2% (Maths low/middle pass) and 1.5% (History high pass).

The facilitating Highers' attainment models were estimated also for those in the top two UCAS points' quartiles for SGI achievement nationally. Whilst comparison of these models with those for the whole population indicated that the magnitude of SIMD effects was reduced for those with an above median SGI performance, comparison of the average marginal effects revealed that there were no significant differences (hence these results are not reported). This is to be expected because individuals enrolled on facilitating Highers are likely to be a more homogenous group, effectively having been filtered at SGI level according to their prior attainment.

2.5.3 Advanced Higher Attainment

Tables 2.15 and 2.16 show the results for Advanced Higher attainment in the Humanities and Languages and Maths and Science groups of facilitating subjects, for all students from 2002 to 2009, for the main variables of interest. Gender effects in humanities and languages subjects are quite varied. Boys were less likely than girls to gain a pass at any level for AH Geography and less likely to achieve a low or middle pass for AH English but were 34% more likely to obtain a high pass in this subject. For AH French and German, boys were significantly more likely than girls to achieve both middle and high passes, being 59% and 73% respectively more likely to achieve a middle pass in these subjects, 82% more likely to achieve a high pass in French and almost twice as likely to achieve a high pass in German. On average, at high pass, these results translated into a respective, 3%, 5% and 7% increase in the probability of obtaining a high pass in English, French and German (Table 2.17). Gender was not significant at all for History, Modern Studies or Spanish. For Maths and science at AH level (Table 2.16), there was a reversal of male/female fortunes. With the exception of low passes in Biology or Physics, where there was no difference, boys were less likely than girls to achieve a given pass level in Maths and the sciences. Average marginal effects (Table 2.18) indicated that being male reduced the probability of achieving a middle or high pass in AH Maths and Science by some 3% - 5%. Prior attainment as measured by individuals' total Higher UCAS points was significant across all subjects at almost all levels [64] with the impact ranging from a 0.2% to 1% increase in the likelihood of obtaining a given pass at AH level for an extra point (Tables 2.15 and 2.16).

[64] The exceptions were: Biology high pass, Spanish middle pass, French/German low and middle pass, where individuals' total Higher UCAS points were insignificant.

Table 2.15

Humanities & Languages Advanced Higher Attainment 2002-2009 : Population

(omitted category: Outcome 1 Fail)

	English	Geography	History	Modern Studies	French	German	Spanish
Outcome 2: Pass low							
Sex (Male=1)	0.789***	0.722**	1.154	0.824	1.199	0.936	0.987
Hpts	1.003***	1.005***	1.004***	1.005***	1.002	1.004	1.007*
SIMD 1	0.777*	0.738	0.712	0.740	0.661	0.482	0.438
SIMD 2	0.738**	0.629*	0.825	0.809	0.840	0.467*	0.657
SIMD 3	0.914	0.937	0.853	0.812	1.291	0.509	0.442
SIMD 4	1.018	0.697*	1.086	1.006	1.123	0.918	1.054
AH cohort size	1.006*	1.002	1.008**	1.002	0.997	0.993	1.004
% pupils SIMD 1 & 2	1.000	0.984***	0.993*	0.996	0.993	0.989	1.006
Outcome 3: Pass middle							
Sex (Male=1)	0.860*	0.566***	1.199	0.824	1.593**	1.726*	1.497
Hpts	1.005***	1.007***	1.005***	1.009***	1.003*	1.005	1.004
SIMD 1	0.641**	0.441*	0.725	0.554**	0.708	0.370	0.532
SIMD 2	0.661***	0.563**	0.769	0.792	0.779	0.630	0.775
SIMD 3	0.771*	0.911	0.713*	0.964	1.000	0.626	0.461
SIMD 4	0.984	0.572***	1.106	1.110	1.095	1.083	0.831
AH cohort size	1.009*	1.003	1.013**	1.007	0.997	1.007	1.017
% pupils SIMD 1 & 2	0.998	0.979***	0.986***	0.992	0.991*	0.986	1.012
Outcome 4: Pass high							
Sex (Male=1)	1.335**	0.393***	1.082	0.917	1.816***	1.930*	0.786
Hpts	1.006***	1.006***	1.007***	1.009***	1.004***	1.008**	1.007*
SIMD 1	0.583*	0.576	0.450*	0.715	0.403*	0.546	0.299
SIMD 2	0.490***	0.547*	0.844	0.613*	0.540*	0.455	0.777
SIMD 3	0.816	0.891	0.655	0.698	0.993	0.535*	0.521
SIMD 4	1.044	0.663	0.943	0.987	1.001	0.643	0.516
AH cohort size	1.013*	1.008	1.012*	1.016*	0.998	0.998	1.011
% pupils SIMD 1 & 2	0.992	0.969***	0.979***	0.992	0.994	0.976*	1.007
Observations	6988	3950	4086	3168	2288	755	460
Log pseudolikelihood	-8454.5615	-4696.8929	-5159.8727	-4055.3316	-3016.8164	-952.98772	-582.57376
Wald chi2(69, Geog 67 Spa 62)	1967.47***	-	7424.28***	6916.21***	4534.07***	4243.83***	-
Pseudo R^2	0.032	0.054	0.067	0.053	0.031	0.077	0.082

Notes:

Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.

SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.

Other independent variables included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects.

Cluster-robust standard errors at the school level (Clusters by subject: English 313, Geography 240, History 265, Modern Studies 182, French 265, German 167, Spanish 72).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.16
Maths & Science Advanced Higher Attainment 2002-2009 : Population

(omitted category: Outcome 1 Fail)

	Maths	Biology	Chemistry	Physics
Outcome 2: Pass low				
SEX (MALE=1)	0.745***	0.957	0.688***	0.916
Hpts	1.003***	1.003***	1.004***	1.004***
SIMD 1	0.727**	0.766*	0.748	0.762
SIMD 2	0.646***	0.831	0.746**	1.019
SIMD 3	0.875	0.970	0.794*	1.018
SIMD 4	0.881	0.901	0.892	0.966
AH cohort size	1.004*	1.004*	1.002	1.002
% pupils SIMD 1 & 2	1.002	0.995*	0.994**	0.994*
Outcome 3: Pass middle				
SEX (MALE=1)	0.679***	0.794**	0.608***	0.786*
Hpts	1.005***	1.004***	1.005***	1.002*
SIMD 1	0.655***	0.612**	0.581**	0.635*
SIMD 2	0.839	0.673**	0.690**	1.188
SIMD 3	0.891	0.701***	0.795*	0.804
SIMD 4	0.866*	0.737***	0.883	0.950
AH cohort size	1.006**	1.007**	1.006**	1.003
% pupils SIMD 1 & 2	1.000	0.993**	0.997	0.989***
Outcome 4: Pass high				
SEX (MALE=1)	0.686***	0.722**	0.612***	0.660***
Hpts	1.004***	1.002	1.007***	1.004***
SIMD 1	0.509***	0.381***	0.342***	0.591*
SIMD 2	0.625***	0.544***	0.572***	0.954
SIMD 3	0.700***	0.681***	0.620***	0.764*
SIMD 4	0.844	0.753**	0.780*	0.824
AH cohort size	1.011**	1.005	1.007**	1.011***
% pupils SIMD 1 & 2	0.999	0.989**	0.990**	0.988**
Observations	11016	7325	7205	5366
Log pseudolikelihood	-14285.186	-9414.3678	-9611.1836	-7252.0872
Wald chi2(69)	668.18***	2275.91***	1781.52***	1157.98***
Pseudo R^2	0.022	0.027	0.025	0.022

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant. SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households. Other independent variables included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects. Cluster-robust standard errors at the school level (Clusters by subject: Maths 329, Biology 310, Chemistry 321, Physics 311). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.17

Humanities & Languages Attainment at Advanced Higher :

Population Average Marginal Effects of Gender & Socio-economic Background

Variable	Grade	English	Geography	History	Modern Studies	French	German	Spanish
Sex (M=1)	Fail	0.030**	0.055***	-0.025	0.028	-0.073**	-0.057	-0.011
	Low pass	-0.051***	0.037*	0.012	-0.019	-0.024	-0.079	-0.016
	Medium pass	-0.008	-0.042**	0.018	-0.017	0.047*	0.067	0.085
	High pass	0.029***	-0.050***	-0.005	0.008	0.051**	0.069*	-0.058
Higher points	Fail	-0.001***	-0.001***	-0.001***	-0.001***	-0.001**	-0.001*	-0.001*
	Low pass	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	Medium pass	0.000***	0.001***	0.000	0.001***	0.000	0.000	0.000
	High pass	0.000***	0.000	0.000***	0.000**	0.000*	0.001*	0.000
SIMD 1	Fail	0.062**	0.060*	0.070**	0.066*	0.099	0.148*	0.144
	Low pass	0.000	0.060	-0.009	0.007	-0.017	-0.050	-0.047
	Medium pass	-0.043	-0.115	0.003	-0.077	0.004	-0.096	0.009
	High pass	-0.019	-0.005	-0.064	0.003	-0.086	-0.003	-0.106
SIMD 2	Fail	0.069***	0.060**	0.036	0.042	0.059	0.126*	0.054
	Low pass	-0.010	-0.008	-0.012	-0.001	0.011	-0.080	-0.049
	Medium pass	-0.029	-0.038	-0.025	-0.002	-0.008	0.004	-0.003
	High pass	-0.030*	-0.014	0.000	-0.039	-0.062	-0.050	-0.001
SIMD 3	Fail	0.028	0.009	0.045*	0.027	-0.022	0.112*	0.127
	Low pass	0.011	0.002	0.014	-0.024	0.053	-0.071	-0.068
	Medium pass	-0.033*	-0.007	-0.031	0.031	-0.018	-0.010	-0.046
	High pass	-0.006	-0.004	-0.028	-0.034	-0.013	-0.032	-0.013
SIMD 4	Fail	-0.002	0.051**	-0.011	-0.006	-0.016	0.021	0.036
	Low pass	0.005	0.006	0.012	-0.008	0.017	0.000	0.061
	Medium pass	-0.006	-0.056**	0.014	0.022	0.008	0.039	-0.001
	High pass	0.003	-0.001	-0.015	-0.008	-0.009	-0.060	-0.095*

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* p < 0.05, ** p < 0.01, *** p < 0.001

Table 2.18

Maths & Science Attainment at Advanced Higher :

Population Average Marginal Effects of Gender & Socio-economic Background

Variable	Grade	Maths	Biology	Chemistry	Physics
Sex	Fail	-0.003*	0.007***	0.006***	0.008***
	Low pass	0.008***	0.017***	0.006**	0.043***
	Medium pass	0.009***	-0.013***	0.006*	0.039***
	High pass	-0.014***	-0.010**	-0.018***	-0.090***
SIMD 1	Fail	0.106***	0.067***	0.046***	0.070***
	Low pass	0.243***	0.132***	0.123***	0.119***
	Medium pass	-0.122***	0.014*	0.075***	0.059***
	High pass	-0.227***	-0.212***	-0.243***	-0.248***
SIMD 2	Fail	0.078***	0.052***	0.033***	0.050***
	Low pass	0.178***	0.101***	0.093***	0.092***
	Medium pass	-0.091***	0.004	0.051***	0.043***
	High pass	-0.165***	-0.157***	-0.177***	-0.184***
SIMD 3	Fail	0.056***	0.033***	0.021***	0.038***
	Low pass	0.131***	0.070***	0.064***	0.064***
	Medium pass	-0.069***	-0.005	0.034***	0.026***
	High pass	-0.118***	-0.099***	-0.120***	-0.129***
SIMD 4	Fail	0.031***	0.019***	0.007*	0.020***
	Low pass	0.077***	0.040***	0.042***	0.037***
	Medium pass	-0.040***	0.003	0.022***	0.013*
	High pass	-0.068***	-0.062***	-0.071***	-0.070***

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: School SGI UCAS points' quartile, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Socio-economic background effects were sparse among the AH Humanities and Languages subjects and where these occurred, primarily in English and Geography, there was no consistent pattern (Table 2.15). By and large, there were no SIMD effects for Modern Studies, German or Spanish at all pass levels, none for History at low pass or French at low and middle pass. Socio-economic background effects were not consistently present across AH Maths and science subjects at low pass but became more prevalent as pass levels rose (Table 2.16). For Maths, Biology and Chemistry at high pass, the established pattern of impact reasserted itself with the likelihood of obtaining this grade falling with individuals' SIMD. Similarly, significant average marginal effects for both subjects were sparse. For Humanities and Languages at AH level, being from SIMD 1 increased the probability of failing most subjects by 6% - 15% (Table 2.17). [65] For those from SIMD 1-3 households, the probability of failing a Maths or Science Advanced Higher was increased by 4% - 11%, while the probability of achieving a high pass was reduced by 3% - 11% (Table 2.18).

Advance Higher cohort size was significant at all pass levels for English, Maths, History and Modern Studies at high pass and intermittently for science (Tables 2.15 and 2.16). Where it was significant, the impact was always positive, generally ranging from a 0.5% to 1% increase in the likelihood of obtaining a given pass (as opposed to failing) for a one pupil increase in cohort size. [66] This cohort size effect at AH level may be indicative of a restricted supply of courses as, often, schools are only able to offer a very narrow range of subjects at this level due to limited resources. In large urban areas, there is often shared access to AH courses across schools with pupils travelling to another school to take a given AH subject not provided at their own school. Such shared access/provision is likely to be impossible in remote rural areas.

Schools' socio-economic composition reduced the likelihood of obtaining any pass grade in AH Geography, History, Biology and Physics and had intermittent/sparse effects on other subjects (Tables 2.15 and 2.16). Its impact ranged from approximately a 0.5% to 3.1% reduction in the likelihood of obtaining a given pass (compared to failing) for a one

[65] The exceptions were French and Spanish.

[66] Exceptions to this were History, Modern Studies and Spanish at high pass where the likelihood of obtaining this grade was 2%, 3% and 5% greater respectively for a one pupil increase in AH cohort size.

percent increase in the proportion of pupils from SIMDs 1 and 2.

As above for Highers, the facilitating Advanced Highers' attainment models were estimated also for those in the top two UCAS points' quartiles for SGI achievement nationally. Comparison of the odds ratios and average marginal effects for these models with those estimated for the whole population indicated that there were no substantive differences between the two sets and hence these results are not reported. Again, this is to be expected; individuals enrolled on facilitating Advanced Highers are likely to be even more homogenous as a group than at Higher, effectively having been filtered twice (at both SGI and Higher) according to their prior attainment.

2.6 Conclusions

Socio-economic-based inequalities in educational attainment in Scotland are long-standing and persistent and have been well-documented in both academic and policy research. The attainment of individuals from disadvantaged backgrounds is significantly and systematically less than that of those from more affluent households. There is evidence to suggest that increasing levels of attainment and participation at age 16 have pushed the critical period for educational inequalities up to age 18. After the various widening access initiatives at the beginning of the 2000s, post-compulsory school participation was seen to rise fastest among low attaining 16-year-olds whilst attainment increased most among those with high age-16 attainment. Gender differences in attainment, in terms of the relative underachievement by boys, have been present since the 1970s and have also been well-documented across the UK. Behind average attainment statistics, there is evidence of perceived traditional patterns of gender performance with males, particularly high attainers, appearing to fare better in Maths and sciences. In general, socio-economic effects have been seen to have a greater impact on educational performance than any gender effects. Some previous research has found individuals' aspirations and motivation to be important determinants of attainment; if these are linked to cultural or social capital, they may in turn contribute to socio-economic effects. Other work has speculated that differential gender attainment may have social and/or genetic determinants in terms of cognitive development.

The policy emphasis is on the reduction of socio-economic rather than gender inequalities. The Scottish Government is committed to reducing the poverty attainment gap in secondary schools and increasing university access for those from the most deprived backgrounds. Progress in closing the poverty attainment gap in secondary schools is to be measured by the difference in the proportions of individuals from SIMDs 1 and 5 (the 20% most deprived households and the 20% least deprived households) achieving one or more qualifications at the different levels. Inequalities in HE entry are to be addressed by introducing access thresholds for all degree programmes in Scotland by 2019, enabling prospective students from the most deprived backgrounds to be assessed against the minimum academic standards and subject knowledge needed for successful completion of a degree programme.

Measuring attainment in terms of the bald number of awards that individuals obtain at given levels, however, is not particularly revealing; secondary school subject choices matter. Previous research has shown that inequalities in HE, in both Scotland and the UK, are most acute on entry and are the result of those from more disadvantaged backgrounds not having the required qualifications. As discussed in Chapter One, some subjects are more important than others for HE entry and/or access to certain degree programmes and employment opportunities. Chapter One demonstrated that individuals from more deprived households tend to study fewer academic subjects, those that facilitate entry to more prestigious universities and/or courses, than those from more affluent households. It is clear that poor secondary school subject choices can adversely affect individuals' life chances and, to the extent that these are socially or culturally driven, provide a mechanism by which educational inequalities can be generated and sustained. To evaluate the need for access thresholds thoroughly, it is attainment *within* facilitating subjects that needs to be examined, once *good* subject choices have been made. This chapter employed multinomial logit analysis to investigate the extent of any attainment gap in Facilitating Subjects, at all levels of qualification in state secondary schools, using SQA administrative data that provide a census of achievement from 2002 to 2009.

In terms of gender effects, at Standard Grade/Intermediate 2, in general, males were found to be less likely to pass subjects at all grades compared to females, but top-performing boys were more likely to pass Maths at all levels and Chemistry and Physics at low and middle pass. There were few gender effects in humanities and languages' attainment at Higher with no clear pattern. At low pass, there were no gender differences except for English with males being less likely to pass this than females. Males were seen to be less likely than females to achieve high passes in English, Geography and German. For Maths and science Highers, however, there were significant, substantial gender effects across all subjects, at all pass grades in favour of males. Males were seen to be some 70% more likely than females to achieve high passes in Maths and Biology and more than twice as likely to obtain high passes in Chemistry and Physics. Gender effects in humanities and languages subjects at Advanced Higher were quite varied and did not conform to any particular pattern. However, gender fortunes were reversed in Maths and science subjects, with males seen to be less likely than females to obtain any given grade.

In common with previous research, socio-economic background effects were found to be greater than gender effects. These were stark at Standard Grade/Intermediate 2, with the likelihood of securing any passing grade in any subject falling dramatically as socio-economic advantage reduces sequentially from SIMD 5 (the least deprived 20% of households) to SIMD 1 (the most deprived 20% of households). The impact was reduced for the top 50% of achievers at low and middle pass for both humanities and languages and Maths and science groups of subjects. For the top 50%, there was a general upwards movement across these attainment grades with, on average, a reduced probability of obtaining a low pass and an increased probability of securing a middle pass. There was, however, no difference in the magnitude of SIMD effects at high pass for the top 50% of achievers, who experienced the reductions in the probability of securing such similar to those for the general population. At Higher, socio-economic effects were found to be much reduced in magnitude at all grades in all subjects compared to Standard Grade/Intermediate 2 qualifications, tending to increase in impact at middle and high grades although not uniformly. Again, in general, where these effects were significant, as pass grades rose the influence of socio-economic effects increased, sequentially reducing the likelihood of obtaining a middle or high pass, the lower an individual's SIMD. The influence of socio-economic background effects largely disappears at Advanced Higher, only showing any consistency at high pass in Maths and Sciences in terms of the established inverse relationship between attainment and an individual's background.

Overall, these findings suggest that once subject choice has been accounted for in terms of the uptake of facilitating subjects after age-16, the impact of socio-economic effects is less. The reduced impact of socio-economic effects at the senior qualifications' level reflects the outcome of a filtering or selection process whereby only higher achieving individuals proceed to take these subjects at Higher and Advanced Higher on the basis of their SGI attainment (securing 'high' Credit passes) and teacher decisions. Thus, it is the level of individuals' SGI pass grades that are of critical importance for securing access to Higher courses in facilitating subjects, rather than simply securing a pass as emphasized in policy targets. While a child's ability/prior attainment is clearly important in terms of securing a low pass in facilitating subjects at Higher, the results suggest that it is not enough to overcome disadvantage to achieve the higher grades that may be required for entry to more prestigious universities and/or courses. This would suggest that, once good subject choices have been made, there is a case for the introduction of access thresholds and a two-tier HE entry system as proposed by the Scottish Government.

This chapter has shown that there are clear gendered patterns of subject choice and attainment in facilitating subjects in Scottish secondary schools. The use of administrative data that provide an effective census of attainment for the period further confirms findings from previous research for both Scotland and the wider UK. Much of this research suggests that gendered subject choice and attainment is probably socially constructed whilst some speculates as to whether there might be a genetic or biological element. The results for Maths and the named sciences at Higher, the crucial qualification level for university entry, are of particular interest given the importance of STEM subjects to income and employment generation in the Scottish economy. The following chapter examines whether there might be a genetic element to STEM subject choice and attainment by investigating the influence of potential exposure to heightened testosterone levels.

Chapter Three

Gender and STEM Subject Choice

3.1 Abstract

Science, Technology, Engineering and Mathematics' knowledge and skills are of increasing importance to the Scottish (and UK) economy in terms of the growth of national income/output and the generation of better-paid employment opportunities. The shortage of suitably qualified individuals in these areas makes it necessary to increase the study of the so-called *STEM* subjects at both tertiary and, importantly, secondary levels of education as the latter feeds the former. STEM subject choice at the level of Higher in Scotland (age-16) is crucial as these are the qualifications that determine both HE and course entry. There are clear, persistent, gendered patterns of STEM subject choice at both SGI (now National 5) and Higher level with, for instance, Biology being far more likely to be studied by females and Physics by males. Differences in educational choices and attainment between females and males have been attributed to social conditioning and gender-biased environments rather than any innate, biological differences with little known about the latter. To help inform policy, it is important to know if *nature* as well as *nurture* factors might be at play too because, if so, the role of socialisation and stereotyping may be exaggerated. This paper explores whether biological factors may have a part to play in the choice of and attainment in STEM subjects in Scottish state secondary schools. Specifically, it examines the impact that the potential exposure to increased levels of testosterone, in-utero, for female twins from mixed-sex twin pairs (the Twin Testosterone Transfer – TTT hypothesis) may have on their STEM subject choices and attainment. Twins and closelyspaced sibling pairs were identified to produce a sub-sample from the SQA national database of candidate results for all state secondary schools for 2002-2009. The closely spaced sibling pairs were used as a control group to attempt to unpack any separate genes and socialisation effects. Logistic and multinomial regression were used to examine subject choice and attainment, respectively, in Maths, Biology, Chemistry, Physics, general Science and Computing. No evidence of any TTT effect on either STEM subject choice or attainment was found although there were clear patterns of gender segregation in subject choice at all levels of qualification. The impact of socio-economic background, however, was very stark. The odds of individuals from disadvantaged backgrounds taking a named science, as opposed to general Science, fell

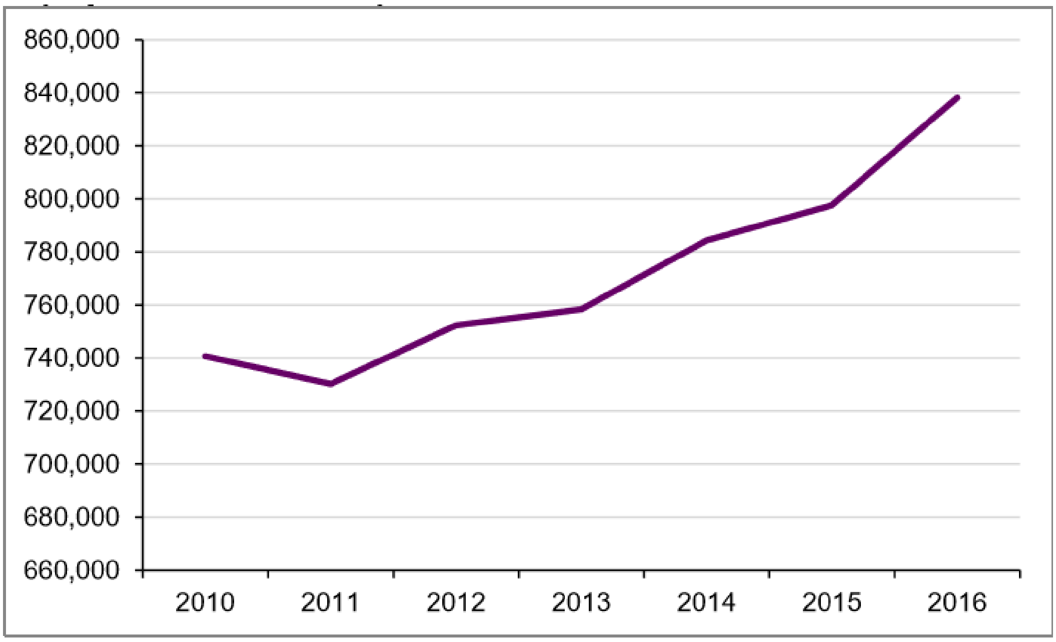
as the level of deprivation rose as did the odds of obtaining middle or high passes in these subjects. Educational choices and general attainment at SGI level were seen to be critical for the uptake of STEM subjects at Higher, indicating that policy interventions to increase STEM uptake need to begin before National 5 level to redress social inclusion and gender imbalance issues. Raising general attainment at National 5 level is, however, only part of the solution with more work needed to understand young people's aspirations and gendered choices.

3.2 Introduction & Policy Context

The Scottish and UK governments are committed to increasing female participation in STEM in schools, universities and the workplace for economic, as well as equity, reasons. Growth in STEM employment in Scotland has been rapid since 2010, rising by 13% (97,500 jobs) as shown in Figure 3.1 below. STEM related employment has been forecast to grow by 4% in Scotland, approximately 42,600 jobs between 2015 and 2027 with much of this growth concentrated in the Central Belt (Scottish Government, 2017c).

Figure 3.1

Growth of Employment in STEM occupations in Scotland, 2010-2016

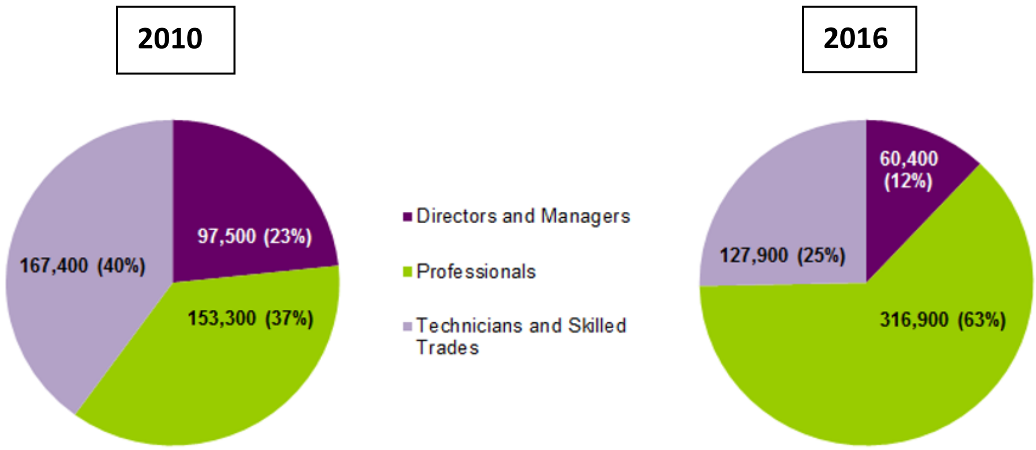


(Source: Scottish Government, 2017c)

Professional STEM occupations have shown the greatest increases in job numbers, particularly in information technology. Professional occupations now account for some 63% of all core STEM jobs (Figure 3.2) and are forecast to continue to grow as a proportion of STEM employment while technician and skilled trades jobs are predicted to decline. At UK level, jobs requiring STEM skills are expected to increase at twice the rate of other occupations over the next decade (UK Commission for Employment & Skills, 2016).

Figure 3.2

Core STEM occupations in STEM industries, 2010 to 2016



(Source: Scottish Government, 2017c)

This growth in STEM jobs is taking place against a background where employers are reporting substantial recruitment difficulties and skills' shortages with, at UK level, some 43% of professional vacancies in science, research, engineering and technology having been identified as *hard to fill* in 2013 (UK Commission for Employment & Skills, 2015). It has been estimated that there is a shortfall of some 40,000 STEM graduates in the UK each year despite recent increases in numbers (Broughton, 2013). This would suggest that the skills' gap is set to worsen unless the number of technically trained individuals in the labour market is increased substantially. From 2020 onwards, it has been estimated that the cost to the UK economy of a shortage in engineering skills alone will be approximately £27 billion per annum (UK Commission for Employment & Skills, 2015). In Scotland, some 73% of women STEM graduates do not pursue careers in STEM fields compared to 48% of male graduates and it has been estimated that doubling the number of highly qualified women in STEM employment could increase Scotland's annual national income by as much as £170m (Royal Society of Edinburgh, 2012).

There is a strong equity case to widen and not just increase STEM participation as higher paid, higher status employment opportunities are open to scientifically literate individuals and citizens should be empowered to 'understand, participate in, and shape scientific developments in society' (Archer et al, 2013, p2). The Scottish Government estimates that those working in the fields of science, technology and engineering, on average, will earn almost 20% more five years after graduation. Full-time, (first degree) STEM graduates from Scottish universities experience a £1,000 per annum premium with an average, annual salary of £23,500 compared to the all degree subject average salary (Scottish Government, 2017f). There are notable gender disparities within STEM occupational groups with men tending to be over-represented in areas such as architecture, engineering and construction, over 80% male, and women over-represented in medicine/dentistry and general health care jobs, approximately 80% female (Scottish Government, 2017c). Under-representation of either gender in the different STEM fields may simultaneously reduce employment/career opportunities and income for individuals, the pool of qualified, talented labour from which employers can recruit and national income. Moreover, it has been shown also that individuals from more disadvantaged socio-economic backgrounds are far less likely to obtain STEM qualifications or employment (Scottish Government, 2017c).

Meeting the Scottish economy's growing demand for professional level STEM skills requires a focus not only on the better retention of female (and male) STEM graduates, but also on growing the STEM skills' base in secondary schools to enable young people in general to apply for STEM courses at university and encouraging the uptake of STEM subjects at this level by, in particular, young women and those from more disadvantaged backgrounds. STEM subject entries and passes for SQA national qualifications for the years directly after the period of this study are shown in Table 3.1. It is difficult to assess changes in participation at the lower qualification levels of SCQF 3-5 as the background is muddied by both demographic change (falling school rolls) and the switch from Standard Grade/Intermediate 2 (SGI) to National 4 and National 5 in this period. Entries for STEM subjects at Higher (H) and Advanced Higher (AH), Levels 6 and 7, were seen to rise between 2010 and 2015 but fell in 2016 by 5%. Between 2010 and 2016, the percentage of passes remained constant at H level but rose at AH level by 17%, masking a fall at both levels in 2016; by 6% at H level and 4% at AH level. Pass rates at Higher fell slightly between 2010 and 2015, down by 1.6%. Falling entries and pass rates at Higher in STEM subjects need to be monitored as this is the crucial stage at which, in the Scottish education system, individuals freely choose the subjects they study and it is these subject choices that determine university entry and future employment prospects.

Table 3.1:
Entries and Pass Rates for National Qualifications in STEM Subjects for Scottish School Pupils, 2010-2016

	2010	2011	2012	2013	2014	2015	2016	% or pp Change 2010/16	% or pp Change 2015/16
SCQF Level 3-5									
Entries	223,423	221,308	222,601	216,227	208,358	205,783	202797	- 9%	-1%
Passes	199,152	198,723	198,393	193,765	165,771	164,174	162,026	-19%	-1%
Pass rate	89.1%	89.8%	89.1%	89.6%	79.6%	79.8%	79.9%	-9.2pp	0.1pp
SCQF Level 6									
Entries	65,652	66,582	66,670	67,115	70,083	71,027	67,363	3%	-5%
Passes	48,554	49,612	50,155	50,052	51,145	51,759	48,741	0%	-6%
Pass rate	74.0%	74.5%	75.2%	74.6%	73.0%	72.9%	72.4%	-1.6pp	-0.5pp
SCQF Level 7									
Entries	10,410	11,143	11,686	11,881	12,099	12,388	11,805	13%	-5%
Passes	7,829	8,574	9,029	9,353	9,206	9,510	9,145	17%	-4%
Pass rate	75.2%	76.9%	77.3%	78.7%	76.1%	76.8%	77.5%	2.3pp	0.7pp

(Source: Scottish Government, 2017c)

Males were seen to dominate the take up of STEM subjects at all levels of qualification in Scottish Schools in these same years; approximately in the ratio 55:45 (male:female) for lower level qualifications and Highers, increasing to 58:42 at Advanced Higher (Scottish Government, 2017c). Female pass rates, however, are generally higher than those for males in STEM subjects at all levels. There are noticeable by-gender splits between subjects with males accounting for nearly 75% of passes in Physics at National level while females account for the majority of Biology passes. Mathematics and Chemistry passes at National and Higher levels are more or less evenly spread between males and females while in Technology subjects, double the number of males leave school with a National level qualification (*ibid*). The attainment gap between those from the least deprived (most affluent) 20% of Scottish households and those from the most deprived 20% is substantial with, for instance, 40% of school leavers from the former achieving at least a pass in Higher Mathematics compared to only 10% from the latter (*ibid*).

At FE level, males accounted for two thirds of STEM enrolments and 73% of credits in 2015/16, suggesting that in addition to signing up for more overtly STEM college courses, they also take more STEM subjects as course components (*ibid*). STEM enrolments at both undergraduate and postgraduate level in Scottish universities accounted for 49% of total enrolments in 2015/16 (114,740 students) with Subjects Allied to Medicine taking the largest share of these at 25% (Table 3.2).

Table 3.2:
Scottish University Enrolments by STEM-related subjects, 2010/11 and 2015/16

	2010/11		**2015/16**		**Change in Enrolments**	
Subject	**Count**	**Share %**	**Count**	**Share %**	**Count**	**%**
Subjects allied to medicine	30,875	28%	29,130	25%	-1,745	-6%
Biological sciences	19,335	18%	21,850	19%	2,515	13%
Engineering & technology	18,245	17%	20,250	18%	2,005	11%
Physical sciences	10,695	10%	11,665	10%	970	9%
Computer science	9,655	9%	10,690	9%	1,035	11%
Medicine & dentistry	7,225	7%	7,655	7%	430	6%
Architecture, building & planning	6,655	6%	5,600	5%	-1,055	-16%
Mathematical sciences	3,765	3%	4.405	4%	640	17%
Agriculture & related subjects	1,575	1%	1,975	2%	400	25%
Veterinary science	1,430	1%	1,520	1%	90	6%
Total	**109,455**	**100%**	**114,740**	**100%**	**5,285**	**5%**

(Source: Scottish Government, 2017c)

Females accounted for the majority of all STEM enrolments (52%) but, as might be expected, there is significant, broadly stable variation by subject. Enrolments in subjects allied to medicine and Veterinary science were predominantly female, 81% and 78% respectively, whist female enrolments in Engineering and Technology and Computer science were just 18% and 20%. Gendered uptake of STEM subjects in schools, particularly at H level, and subsequently at FE/HE is reproduced in the pattern of gender segregation in the labour market (see Kim & Kim, 2003 for a discussion).

The Scottish Government's STEM Education and Training Strategy aims to redress the gender and socio-economic imbalance present in STEM education and employment, developing increased STEM capability and skills in learners across the educational spectrum from early years/primary to tertiary education. Several key performance indicators have been identified. [67] In terms of secondary school education, together with a general increase in the promotion of STEM in deprived and/or rural communities, the specific aims are:

- 10% increase in passes in Mathematics at SCQF level 5 by 2022
- Expansion of Foundation Apprenticeship opportunities across all Scottish Secondary Schools
- Reduction in the attainment gap between school leavers from the least deprived and most deprived SIMD quintiles with 1 or more awards in STEM subjects, at SCQF level 6 (Higher) or better, to 31 percentage points by 2020 and 25 points by 2022
- Improvement of the gender balance in attainment in key STEM subjects at Higher by 2022, increasing the number of females passing Physics by 15% and Computing by 20%

Gender imbalance in the uptake of Physics by females would seem to present a particular problem England and Wales too. Whilst numbers studying A-level Physics increased by 15% between 2010 and 2016, only 20% of candidates were female as has been the case for the last thirty years (Institute of Physics, 2018). The Institute of Physics (*ibid*) found that, compared to boys, girls in mixed state schools were less likely to progress to 'A' Level Physics than those in single-sex, independent schools. When Physics was one of girls' top four GCSE grades, only 8% continued to A-Level compared to 25% for Chemistry and 32% for Biology when these were among top four grades. Almost 16% of girls continued to A-Level Biology when it was not one of their top four GCSEs; ie almost double the proportion that continue with Physics when it is one of their top four grades. The Institute states:

> "We believe that although girls and boys currently choose subjects differently, there is no evidence to suggest there are any intrinsic differences in preference, or in ability, which are reason for them to do this." (*ibid*, p5)

[67] https://beta.gov.scot/policies/science-and-research/stem-education-training/ accessed 19/06/18.

Simply increasing science provision in schools would not seem to be the answer, De Philippis (2016) found that increased science provision in English secondary schools, whilst equally likely to be accessed by girls, led to greater enrolment of males on STEM subjects at university but not females. Interim findings from the ESRC-funded ASPIRES Project investigating children's science and careers aspirations between ages 10-14, found that the three most important influences on these were firstly family, followed by hobbies and interests outside school and then school itself (Archer et al., 2013). [68] Family influence tended to result in aspiring to the same job as a family member or close family friend. Hobbies and interests produced aspirations for sports or arts related careers. School influence, in addition to developing interests and aptitudes in relation to, for instance, Science, resulted in a sizeable proportion of children expressing interest in becoming a teacher. Three main reasons were identified as to why science careers are not popular: a lack of science capital, an image of science as being only for the brainy, and the perception of science as *masculine*. Science capital was rooted in family socio-economic background with privileged children being found more likely to conceive of professional careers in medicine or science as they had immediate role models. Long-established, gendered career aspirations were found to still be prevalent with 12-13 year-old girls more likely to conceive of careers in arts and boys more likely to profess interest in engineering careers. Interestingly, researchers found that most girls with science aspirations regarded themselves as not being "girly" (Archer et al., 2013, p16).

Given the importance of STEM to the economy and for individuals' employment opportunities, and its consequent promotion by policy makers, this paper attempts to examine whether less "girly" girls might be attracted to choose STEM subjects (Maths, Biology, Chemistry, Physics, Computing) in secondary school as opposed to choices being purely environmentally determined. Specifically, it explores whether subject choice might be related to individuals' genetic make-up by exploiting the *Twin Testosterone Transfer* hypothesis; whereby the female of a mixed twin pair may be subject to heightened levels of testosterone in utero. The analysis is again undertaken for state secondary schools in Scotland using administrative data: SQA's candidate database of qualifications' results

[68] The findings were based on the analysis of the responses of some 5,000+ children aged 10-13 to open-ended questions about their carer aspirations and influences (*ibid*).

from 2002-2009. The following section reviews the educational attainment literature, drawing from work in the fields of Behavioural Genetics, Psychology and Sociology in addition to Economics. This is followed by a discussion of the data and methodology used. The results of the analyses are then presented and discussed followed by a conclusion.

3.3 Nature *and* Nurture – a Review of the Literature

Family, neighbourhood and school are often singled out as the three most important factors that influence a child's educational achievement (Leckie et al., 2010). Twin studies have often been used to control for the influence of family which is twofold, providing both the genetic make-up individuals inherit from their parents and the domestic environment in which they grow up. If monozygotic (identical) twins (who share 100% of their genes) can be identified, this enables the family genes' effect to be separated from the family environment effect. Neighbourhood *norms* may impact on educational achievement and uptake; the educational aspirations of children may be lower in more deprived communities where, for instance, adults have low levels of qualifications and/or low- or unskilled jobs. As government policy interventions aiming to improve socio-economic outcomes occur in all three areas, identifying the relative contributions of family, neighbourhood and school is important for the effective formation and targeting of educational and social policy.

In their overview of the literature, Rasbash et al. (2010) point out that these three factors tend to be studied separately. They found that *Value-added* studies of school effectiveness, usually estimated with multilevel models, suggested schools account for up to 20% of the variation in educational achievement with the remaining 80% being attributable to the children themselves. Studies examining the influence of primary schools have found their *carryover* effects to be large (but generally not as great as secondary school effects) while those that have controlled for neighbourhood and/or local authority, have found the effects of these to be small in relation to school effects (*ibid*). Family studies have shown that the IQ and educational achievement of siblings tends to be similar; this may be the result of share genes – *nature* or growing up in the same family – *nurture* (Leckie et al., 2010). Population genetics (multilevel) models have shown, typically, approximately 25% of the variance in children's IQ scores to be attributable to non-genetic, environmental factors that make siblings similar (Rasbash et al., 2010) Such models, however, do not distinguish between environmental influences internal to the family (e.g. parents reading to children) and the external environmental influences of school and neighbourhood.

Rasbash et al. (2010) attempt to separate family and external environmental influences, using derived twin data from the National Pupil Database (NPD) [69] to ascertain: (i) how much of the (80%) variance in educational attainment attributed to the child in school effectiveness studies is the result of family factors and, in turn, (ii) how much of the variance usually attributed to the family is actually due to shared school and neighbourhood environmental influences outside the family. They followed a cohort of 500,000 children through secondary school, analysing information on their academic attainment and schools viz: age 11 Key Stage 2 test scores (end of primary school), age 16 GCSE examination results (end of secondary school), the primary and secondary schools where the children sat these, the schools' LEAs and the neighbourhoods where the children lived. As family information is not present in NPD data, twins were identified through shared dates of birth, postcodes, ethnicity and eligibility for free school meals. Using cross-classified multilevel models, the variance in children's secondary school, academic progress was decomposed into effects attributable to six influences: LEAs, secondary schools, neighbourhoods, primary schools, families and pupils. It was found that 40% of the total unexplained variation in attainment occurs at the level of the family; that is a composite family effect including both shared genes and the influence of immediate family environment as it was not possible to distinguish monozygotic (MZ) and (fraternal) dizygotic (DZ) twins. Of the remaining variation, 38% was at the level of the individual child, 10% and 9% at secondary and primary school level respectively and just 3% at neighbourhood and LEA level. This implies that a total of 22% of the variance in secondary school academic progress is attributable to wider, shared environments outside the family and perhaps indicating some role for in-school peer effects.

3.3.1 Peer Effects

Peer effects on performance in the national high school entrance examination in Taiwan, between 2002 and 2004, were estimated by Chou et al. (2015) using data on twins. A

[69] The NPD provides a census of children in the English state school system.

number of models were estimated: OLS, twin fixed effect and instrumental variables (IV) to take account of endogeneity, conditional quantile regressions, twin fixed effect and IV quantile regressions to ascertain the strength of peer effects at different points on the examination score distribution. These models were used to estimate the influence of peer effects on the entrance examination score in five different subjects: Chinese, English, Maths, Natural Sciences, Social Sciences. Peer effects were measured by the average subject score for an individual's peer group less the individual's own score. Two instruments were used for the IV specifications: (i) average peer group scores for all the other individual subjects (less the individual's score), (ii) the summation of the average peer group scores for all the other individual subjects. Positive, statistically significant peer effects were found in the case of all model specifications. The Fixed effects and IV approaches for both the mean and quantile regressions yielded similar, reduced peer effects estimates compared to the pooled OLS and standard quantile regression models. Results for the same-sex twin samples indicated that a one standard deviation higher mean examination score for peers was associated with a 0.19-0.56 standard deviation increase in an individual's score. Whilst controlling for fixed effects was seen to reduce selection bias in general, the authors found that the estimation of a twin fixed effects model for mixed-sex twins produced similar results to the OLS model (with no fixed effects) for the whole sample. They conclude that these results indicate the persistence of non-random selection in twins' studies particularly with respect to mixed-sex twins. Peer effects were seen to matter most in the middle of the exam score distribution. For both all-boy and all-girl twins, the impact of peer effects in the fixed effect quantile regressions generally exhibited an inverse U shape for the different subjects, with the largest effects occurring at the 0.25, 0.5 and 0.75 and smaller effects in the tails of the distribution.

As siblings share both the family and the neighbourhood environment they grow up in, measures of their similarity include both influences. Nicoletti & Rabe (2010) suggest that by looking at the similarity between siblings and between unrelated neighbours, it is possible to put bounds on the potential magnitude of family and neighbourhood effects. They examine test results at the end of primary school (age 11) and the end of compulsory secondary schooling (age 16) in England, using the NPD. Sibling differences in attainment outcomes are analysed by sibling gender combination, age difference, school starting age difference and genetic similarity, comparing monozygotic (identical), dizygotic (fraternal) twins and non-twin siblings. A variance decomposition approach is used to try to establish bounds on the size of family and neighbourhood effects. Correlations between siblings' and between unrelated neighbours' total GCSE points' scores for English, Mathematics and Science are estimated and their end of primary school test scores in the same subjects. They argued that the neighbourhood correlation in these test outcomes provides an upper bound for the importance of neighbourhood factors in explaining attainment. This is because the neighbour correlation will also include within neighbourhood, positively correlated family traits as a result of sorting/selection. A lower bound for the importance of the family effect is given by subtracting the correlation between neighbourhood children from the sibling correlation. It is suggested that this lower bound for the family effect can be interpreted as a relative family effect. This relative family effect measures the part of the family effect arising from a family having characteristics that differ from those of other families living in the same neighbourhood. Their estimates indicated that deviations of families' characteristics from observed neighbourhood mean family characteristics account for 47% of the deviation in pupils' age-11 and 48% in pupils' age-16 attainments at the lower bound. Shared neighbourhood backgrounds were found to account for at most 10-15% of the differences in educational attainment between pupils. The influence of living in an urban neighbourhood was found to be slightly larger than living in a rural one; it was suggested that this might reflect the potential for greater interaction between individuals in more highly populated areas. Similarity in attainment was seen to be higher between same gender siblings than between siblings of differing gender and was found to be higher at eleven than at sixteen. The similarity in educational outcomes for identical twins was 0.25-0.35 points higher than for fraternal twins, presumably, the authors suggest, because identical twins share exactly the same genes.

3.3.2 Ability & Heritability

Whilst twin studies in Economics have been concerned in the main to estimate the returns to education (for instance Bonjour et al., 2002, Card, 1999) there is a large Behavioural Genetics' literature that focuses on educational outcomes using twin methods. The field of Behavioural Genetics is concerned with identifying the genetic and environmental determinants of individual differences in complex traits [70] throughout individuals' development (Haworth et al., 2011). The focus is on the statistical analysis of variation in genes and environments and their interaction in relation to variation in outcomes. The received wisdom is that educational achievement is influenced by environmental factors such as schools, teachers, parents, peer groups, whereas ability is innate. Considerable genetic research, however, suggests that differences in educational achievement are substantially heritable (i.e. genetically determined), and, in early school years, more heritable than differences in intelligence (Krapohl et al., 2014). In a study that examined the heritability of a composite GCSE test score for English, Maths and Science for 6,653 pairs of twins in the UK Twins Early Development Study (TEDS), Shakeshaft et al. (2013) found high heritability (58%) for the GCSE results. Krapohl et al. (2014), hypothesizing that heritability in educational attainment is explained by more than just intelligence, also examined: self-efficacy, personality, well-being, parent-rated behaviour problems, child-rated behaviour problems, health, perceived school environment, perceived home environment. Using multivariate genetic analysis, they found that these broader domains of behaviour, together with intelligence, jointly explained 75% of the heritability of the GCSE score. [71] GCSE mean score gender differences in favour of females were found but these were very small (males 8.86, females 8.96) although statistically significant due to the very large sample size. Less than one percent of the variance was found to be accounted for by sex, zygosity, and their interaction. Previous studies have found little evidence of gender differences in genetic and environmental estimates of educational attainment (Shakeshaft et al., 2013, Johnson et al., 2006 in Krapohl et al., 2014). Environmental influences, Krapohl et al. (2014) argue, can only be disentangled from genetic influences in genetically sensitive designs such as the twin method. They found

[70] The inheritance of complex traits is polygenic rather than monogenic (i.e. they cannot be explained by a single isolated gene). They are characterised by a continuous range of variation and influenced by both environmental and genetic factors. Height, circadian rhythms and many diseases such as diabetes and Parkinson's are examples of complex traits (https://en.wikipedia.org/wiki/Complex_traits , accessed 06/07/18)

[71] Intelligence was found to explain more of the heritability of GCSE educational achievement at age 16 than any other behavioural trait. The other traits together accounted for nearly as much GCSE heritability as intelligence alone. Collectively, all cognitive and noncognitive predictors accounted for 75% of the heritability of GCSE scores (*ibid*).

that shared environmental influence explained 26% of the variance in GCSE performance and speculate that family and school are the likely sources of influence.

The Krapohl et al. (*ibid*) research is controversial not only because achievement is thought to be driven by the environmental influences of home and school but because the other behavioural traits that contribute to educational achievement are presumed to be environmentally determined also. Their results suggest the opposite, showing that genetic influence is greater for achievement than for intelligence, and that the other behavioural traits thought to impact on educational achievement do so largely for genetic reasons. They suggest that:

> "….. children differ ….. in how easily they learn and perform at the examinations … not just because of differences in intelligence, but because of a whole package of genetically related characteristics including self-efficacy, personality, and behaviour problems, as well as intelligence." (*ibid*, p15276)

but posit that these phenotypic correlations between such traits and educational achievement can be mediated environmentally though policy interventions.

To measure environmental influence in terms of the value added by schools, it is necessary to identify the contribution of both genetics and family environment to academic achievement. Haworth et al (2011) use a twin method to investigate whether better measures of school 'added value' can be obtained from academic achievement measures that have been corrected for previous attainment in terms of both general cognitive ability and previous school performance. Achievement was measured by a combination of year-long teacher assessments and formal test scores for a sample of 4,000 pairs of 12 year-old twins from the UK TEDS. Uncorrected raw achievement was shown to exhibit *moderate* heritability at approximately 50% and shared environmental influences of 25%. Achievement independent of previous achievement demonstrated just as much genetic influence as raw achievement, remaining consistent at 50% in the indices of school value-added. The implication of this is that half of the variance in achievement occurs as a result of environmental differences between children. Shared environment that contributes to sibling similarity (family, school), however, appeared to be less important at only 12% for the corrected achievement scores. The authors suggest that controlling for ability and previous achievement removed much of the shared environmental influence but not the non-shared environmental influence, increasing the proportion of the corrected achievement scores explained by the latter and warranting further investigation of such. Given the clear pervasiveness of genetic influence, they argue that the current move towards individualised education and an active view of learning is to be welcomed, whereby:

> "……children select, modify and create their own education in part on the basis of their genetic propensities." (*ibid,* p9)

Shared environmental impact could emanate from assortative mating (Evans & Martin, 2000), whereby mate selection depends on trait similarity between spouses in, for instance, educational achievement and intelligence. It is suggested by Vinkhuyzen et al. (2012), that if assortative mating is not controlled for, estimates of the relative magnitude of genetic and environmental factors may be biased upwards. They use an extended twin-family framework to model the effects of assortative mating on the variance components' estimates of intelligence. Psychometric IQ data for a sample of 1314 adult twins (monozygotic and dizygotic), their siblings, the partners of the twins and siblings, and either the parents or the adult offspring of the twins and siblings were used. Both phenotypic (attraction to certain characteristics or traits) and social homogamy (attraction to those from a similar social background) assortment processes were modelled. The phenotypic assortment (PA) model was preferred, providing a better data fit than the social homogamy model for general, verbal and performance intelligence, suggesting that assortment for intelligence is the result of mate selection on similarity in intelligence rather than social background. Contradictory single PA model results,[72] probably caused by OVB, led the authors to suggest two alternative models: (i) controlling for negative cultural transmission (CT) but not genetic dominance [73], (ii) controlling for

[72] In the original PA model, genetic dominance and negative CT were confounded.

[73] Genetic dominance refers to the extent to which the effects of alleles (gene variants) are not additives but interact with each other with one being *dominant*. Dominance in genetics is a relationship between alleles of one gene, in which the effect on phenotype of one allele masks the contribution of a second allele at the same locus. The first allele is dominant and the second allele is recessive (https://en.wikipedia.org/wiki/Dominance_(genetics), accessed 9 May, 2018)

genetic dominance but not negative CT. Additive genetic variance, variance explained by assortative mating and non-shared environmental variance were included in both models. The latter PA model, without CT was preferred suggesting that the variance of intelligence in adulthood to be attributable to: non-shared environmental (18%), additive genetic factors (44%), non-additive genetic factors (27%), phenotypic assortment (11%). The model provided similar results for the verbal intelligence and performance intelligence subscales.

These results suggest that the large impact of additive genetic factors on the heritability of intelligence found in many studies, may arise from genetic dominance and positive assortative mating. Shared environmental influences disappear when the children leave the parental home and the heritability of intelligence has been shown to increase with age (*ibid*). The authors speculate that genetic dominance, shown in their findings to explain part of the variance in intelligence in adulthood, may be masked by larger shared environmental or CT effects in childhood. They assert that the non-additive nature of genetic influences should be included in future genome-wide association studies for intelligence. Additionally, they suggest that childhood shared environmental effects may be overestimated as a resulted of not controlling for assortative mating.

The study has several limitations as pointed out by the authors. Implicitly alluding to the endogeneity of mating behaviour, they reiterate that this was assumed to be either based *strictly* on phenotypic assortment (determined solely by phenotypic similarities) or social homogamy alone (determined by environmental similarities only). It is likely, however, that it is driven by both processes. Furthermore, it is suggested that social homogamy could be the result of genetic stratification between populations and that it is possible also that PA could be environmentally driven if the trait in question is not genetically determined. It is pointed out also that the determinants of assortative mating may have changed over time with similarity in social status perhaps being more important in the first half of the 20th century and phenotypic assortment more influential in the second with increasing urbanization and gender equality in educational opportunities. It has been suggested also that genes and non-shared environmental factors (for instance, education and profession) interact to determine intelligence (Haworth et al., 2010 in Vinkhuyzen et al., 2012) and that ignoring such interactions can lead to the estimates of genetic effects on intelligence being biased upward (Purcell 2002 in Vinkhuyzen et al., 2012).

Ayorech et al. (2017) examined genetic influence on intergenerational attainment assessing the correlation between parent and offspring achievements. They used twin and genomic analyses based on 6,105 twin pairs and 5,825 unrelated individuals taken from the twin sample. Their findings suggested that genetics accounted for nearly half the variation in intergenerational educational attainment. A genome-wide polygenic score (GPS) was calculated for years of education; a genetic risk score, based on the variation in multiple genetic loci and their associated weights.[74] They found the highest (0.43) and lowest (-0.19) GPS means occurred respectively within *stably* educated and uneducated families respectively; that is, individuals who had taken 'A' Levels and had a university-educated parent and those who had not taken 'A' Levels and had no university-educated parent. The mean GPS for *upwardly* and *downwardly* mobile children (those who had taken 'A' levels but did not have a university educated parent and those who had not sat 'A' levels but did have a parent that went to university), fell in the middle of the range (0.05 and 0.28 respectively). The authors suggest that genetic influences on intergenerational educational attainment provide an index of equality of educational opportunity; essentially implying that the greater the genetic impact, the more equal is the educational system in terms of access and allocation. This is in keeping with Kovas et al. (2013), who postulate that the broad equality of opportunity in the relatively homogenous education sytems of the West reduces environmental variation, differences in educational attainment between children, therefore, reflect genetic differences to a greater extent. [75]

Twin studies of educational attainment emanate generally from the UK or the US, however, several Dutch studies have produced similar findings in terms of heritablilty. For instance, Calvin et al. (2012) found genetic effects to be a significant cause of variation in both language (accounting for 43-74%) and arithmetic (36-73%) achievement at ages 8,

[74] A GPS serves as the best prediction for a trait (in this case years of education) that can be made when taking into account variation in multiple genetic characteristics (https://en.wikipedia.org/wiki/Polygenic_score: accessed 27 April, 2018).

[75] As indicated in Chapter Two, research on intergenerational mobility in the UK in 1980s and 1990s (Blanden et al., 2003), suggests that there has been a sharp rise in tertiary educational inequality.

10 and 12 for Dutch primary school children. These findings may be underestimated as the resemblance between monozygotic and dizygotic twins was calculated from the proportion of same-sex and opposite-sex twins. Primary school, national test scores for 6-12 year olds in arithmetic, reading and comprehension, spelling and the (end of primary school) educational achievement test were examined by de Zeeuwa et al. (2016), for just over 7,000 twins from The Netherlands Twin Register (NTR). Significant mean gender differences were found with, on average, boys performing better in arithmetic and girls in reading comprehension and spelling tests. The end of primary school educational achievement test revealed similar gender differences with boys, on average, scoring higher in arithmetic, study skills, science and social studies while girls had better results for language. Estimated twin correlations revealed all MZ correlations to be higher than DZ correlations, implying the existence of additive genetic effects, given that MZ twins are genetically identical. DZ twins share 50% of their genes, therefore, it might be expected that, *ceteris paribus*, DZ correlations would be approximately half the magnitude of MZ correlations. The DZ correlations were sometimes greater than half the MZ correlations, indicating the existence of common environmental effects. The genetic modelling results revealed small gender differences in the estimates of the variance components, but, in general, these were insignificant. Making the distinction between quantitative and qualitative gender differences [76], they conclude that there was no evidence for either, indicating that the impact of genes and the environment is broadly similar for both genders and that the genes that influence educational attainment are the same for both (*ibid*).

For the national, in primary school tests, genetic effects were seen to make the largest contribution to individual differences in educational achievement accounting for variance between: arithmetic - 60-74%, word reading - 72-82%, reading comprehension - 54-64%, spelling - 33-70%. Common environmental effects had negligible influences in general except for spelling (where impact ranged from 0-29%). Unique environmental effects explained the remaining variance, ranging from: arithmetic - 26-34%, word reading - 11-29%, reading comprehension - 32-35%, spelling - 30-39%. The results for the end-of-

[76] Quantitative gender differences occur if one gender is affected to a greater extent by the same genetic or environmental effects while qualitative gender differences are present if different genetic or environmental effects influence boys and girls (*ibid*).

primary school educational achievement test revealed similar patterns to the in-primary test scores with gene effects again making the largest contribution to the variation (74%). However, this test includes Science and social studies which were found to be less heritable (at 56%) compared to arithmetic (68%) and language (67%) and exhibited a larger common environmental influence (21%) compared to the latter (5% for arithmetic and 10% for language). Other studies also report lower heritability for science performance (Haworth, et al., 2009, Haworth et al., 2008), de Zeeuwa et al. suggest that an explanation for this might lie in the greater curriculum and lesson heterogeneity for science and humanities subjects whereas the teaching of language, reading and maths is more standardized. Greater heterogeneity in educational provision increases environmental differences and, ergo, the proportion of the differences between children explained by genes will fall (Heath et al., 1985).

A systematic review of 61 studies addressing the heritability of educational achievement in primary school is provided by de Zeeuwa et al. (2015), together with a meta-analysis of the correlations within (5530) MZ and (7084) DZ twin pairs for general educational achievement and attainment in specific subjects. The studies were weighted by sample size and a variance decomposition model was used to estimate the influence of genetic and shared environmental effects on attainment. Structural equation modelling allowed a maximum likelihood estimate of heritability to be obtained and tested for equality across all the studies. Heritability estimates in different subjects were found to vary widely across the reviewed studies, viz: reading (0.10–0.94), reading comprehension (0.32–0.87), mathematics (0.04–0.75), spelling (0.33–0.84), language (0.21–0.81), science (0.32–0.64) and general educational achievement (0.27–0.57). Environmental effects were found to vary considerably also: reading (0.00–0.74), reading comprehension (0.00–0.50), mathematics (0.00–0.81), spelling (0.00–0.46), language (0.10–0.25), science (0.08–0.39) and general educational achievement (0.08–0.67). The authors speculate that the wide range of estimates may result from: large differences in sample sizes, different countries and/or age groups, variation in measurement instruments. Studies providing separate estimates for the heritability in boys and girls did not report any gender differences (Harlaar et al., 2005, Kovas et al., 2007, Petrill & Thompson, 1994, Reynolds et al., 1996 all in de Zeeuwa et al., 2015, Haworth et al., 2009, de Zeeuwa et al., 2016). The number of studies included in the specific subjects' meta-analysis was small, ranging from 11 for reading to three for maths and only two for general achievement (there were none for science). The estimates revealed genetic effects of: 73% - reading, 49% - reading comprehension, 57% - mathematics, 64% - language, 44% - spelling, 66% - general educational achievement. Estimates of shared environmental effects ranged from 10% for reading and mathematics to 15% for language, with spelling providing a relative outlier at 23%. Equality of the size of the heritability estimate was found for reading and general educational achievement only with results for the other subject areas displaying heterogeneity of impact.

In the main, the meta-analysis studies came from the UK, USA and the Netherlands. The cross-country heritability of reading was found to be consistently high (USA: 69%; UK: 76%; NL: 66%) whilst there was considerable variation in the other subjects. However, heritability was seen to be generally high for all subjects in the Netherlands, ranging from 64% in reading comprehension to 71% in Maths, leading the authors to suggest that differences between countries might reflect differences in educational opportunities. The Dutch education system provides a relatively homogenous educational environment as both public and private schools are state funded and have to comply to the same standards. It is argued that these very similar school environments will reveal differences in attainment between children to be the result of genetic differences to a much greater extent (Kovas, 2013, Heath et al., 1985 in de Zeeuwa et al., 2015). As has been noted in Chapter One, in Scotland, the quality of provision has been seen to vary little from school to school but attainment varies widely; with schools in more affluent areas achieving better levels of attainment than those in poorer areas (OECD, 2007) indicating that other factors are at play.

Overall high heritability of educational achievement, de Zeeuwa et al. (2015) conclude, means that innate individual differences in ability between children will come to the fore; levels of educational attainment can be raised by changes in the environment but the variation between individual children will be largely the result of genetics. The authors speculate further that differences in heritability could reflect differences in income inequality which is larger in the UK and USA compared to the Netherlands (OECD, 2011). There is some inconclusive evidence to suggest that socio-economic status may exert a moderating effect on general cognitive ability. It is postulated that children from more affluent families are likely to have greater opportunities to realize their genetic differences. A number of US studies have found larger heritability of general cognitive ability among middle- and upper-class children while environmental effects were seen to have a larger influence on children from lower income families (Scarr-Salapatek, 1971, Turkheimer et al., 2003; both in de Zeeuwa et al., 2015). A UK study (Hanscombe et al., 2012), however contradicts this, concluding that while variation due to shared environmental was larger in children from more deprived backgrounds, genetic influence on general cognitive ability was equal in children from low and high-income families. If heritability of educational achievement is mitigated by socio-economic status, then reducing the percentage of children from disadvantaged backgrounds will result in higher heritability of educational attainment.

It should be noted that whilst many Behavioural Genetics' studies find educational attainment to be moderately-highly heritable at approximately 40-70% (in the above studies) and correlated with other heritable characteristics - cognitive function, personality traits related to persistence and self-discipline - there has been no discovery of equivalent magnitude of any genetic variants associated with such traits. Rietveld et al.'s (2013) meta-analysis, genome-wide association study of educational attainment found significant but very small effects for three independent single nucleotide polymorphisms (DNA building block differences)[77] on years of schooling and college/degree completion for a replicated sample of 25,490 individuals. These effects equated to approximately 1 month of (extra) schooling per allele (given gene variant) and a 1.8% difference in the frequency of college/degree completion (R^2 of approximately 0.02%). They state that:

> " …. the genetic architecture of complex behavioral traits is far more diffuse than that of complex physical traits. " (*ibid*, p1469).

It is observed that existing claims of "candidate gene" associations with complex social-science traits such as educational attainment have shown great variation in effect size

[77] Single nucleotide polymorphisms (SNPs), are the most common type of genetic variation among individuals. Each SNP represents a difference in a single DNA building block or *nucleotide*. (https://ghr.nlm.nih.gov/primer/genomicresearch/snp, accessed 8 May, 2018).

with much larger (but still relatively small) R^2 values of 4-6%. They suggest, therefore, that their meta-analysis estimate of 0.02% should serve as a benchmark for evaluating the plausibility of findings in respect of the genetic architecture of educational attainment and similar complex social-science phenotypes.

3.3.3 Gender

Analysis of the 2003 Programme for International Student Assessment (PISA) test scores for Maths and Reading for 15-year olds in 41 industrialised countries by Machin and Pekkarinen (2008) reaffirms widely found gender differences. In general, boys were seen to outperform girls in Maths while girls outperformed boys in reading. Additionally, they found evidence to support the view that boys' educational performance is characterised by greater variance than girls'. In the majority of the countries, boys' educational performance exhibited higher variance than that of girls' in *both* subject areas, [78] driven by different compositional effects. The increased presence of boys in the upper part of the test score distribution drives the higher boy-girl Maths' variance ratio, whilst their greater prevalence at the bottom end of the reading test score distribution drives the higher boy-girl variance in this subject. They point out that Higher variability among boys characterises performance in reading and maths tests across the world but note that the difference in variance is higher in countries with higher levels of test score performance

While traditional, significant mean gender differences are commonly found, de Zeeuwa et al.'s (2015) systematic review (as noted above) found no gender differences in the heritability of educational attainment. Deary et al. (2007) found age 11 general intelligence scores to be highly correlated (r >0.8) with (age 16) GCSE exam performance, using a large, representative, longitudinal survey of more than 70,000 English schoolchildren. While they found no differences in the age-11 general intelligence of boys and girls, with the exception of Physics, girls' GCSE performance was significantly better than boys in all other subjects. [79] Another large study (Kovas et al., 2007) that used teacher ratings to ascertain pupils' language, reading, mathematics and science

[78] Boys' educational performance was characterised by higher variance in 37 out of the 41 countries for Maths and 35 out of the 41 countries for Reading (*ibid*).
[79] Results for 25 academic subjects were examined and girls' better performance was not found not to be the result of better verbal ability. (*ibid*).

attainment found no gender related differences between children. Measuring the reading level of children by means of a standardized test, Harlaar et al. (2005) also found an absence of gender differences. General cognitive ability is regarded as the most important predictor of educational achievement (Deary et al., 2007), explaining approximately 50% of the variation (Frey & Detterman, 2004), with no apparent differences in this, mean gender differences in attainment remain unexplained. It is suggested by de Zeeuwa et al. (2016) that mean gender differences in educational attainment might be the result of differences in the determinants of educational performance between the sexes, although the results of their study (in accordance with Kovas et al., 2007) mitigate against

> "… a difference in the genetic architecture of educational achievement as an explanation for the mean differences between boys and girls." (de Zeeuwa et al., 2016, p9)

Testosterone is responsible for the development of male sex organs, dimorphic physical characteristics and behavioural differences; 'wiring the brain with masculine behavioural traits in preferences, personality and temperament' (Jordan-Young, 2010). Exposure to testosterone has been linked to a variety of traits, among which are language development, spatial ability and sexual behaviour and reduced testosterone levels in utero have been associated with the demasculinization or feminisation of these traits (Tapp et al., 2011). The twin testosterone transfer (TTT) hypothesis posits that testosterone from a developing male fetus may be transferred in utero to the other developing fetus(es). Where the co-twin is female, it is argued that testosterone transfer can result in the masculinization of behaviour, cognition and morphology and, for a male co-twin heightened masculinization. [80] Females with male co-twins have been shown to exhibit a more masculine brain structure and volume (Cohen-Bendahan et al., 2004 and Peper et al., 2009 cited in Gielen & Zwiers, 2018). Gielen and Zwiers (2018) investigate whether prenatal testosterone might explain gender differences in primary school educational performance using administrative data for a large sample of Dutch twins. The premise of TTT is used as the basis to examine the impact of the hormone as a potential explanation for gender differences on three measures of educational performance:

[80] https://en.wikipedia.org/wiki/Prenatal_testosterone_transfer, accessed 23/05/2018

overall ability, reading, maths. Given the existence of TTT, individuals with male co-twins would be subjected to higher levels of testosterone in utero, thereby providing an exogenous source of variation in testosterone levels as the sex of a co-twin is randomly determined.

The potential influence of growing up with a same or opposite sex sibling is controlled for too, since, to establish causal effect, a co-twin's sex must not be related to educational outcomes other than through TTT. This is unlikely to be the case as growing up with a brother is different from growing up with a sister and could affect educational outcomes. Comparing the outcomes for twins with male co-twins with those for twins who have female co-twins will capture the effect of both prenatal testosterone and that of growing up with a same- or opposite-sex sibling. They created a control group of closely spaced singletons (CSS), born within one year of each other,[81] to isolate the effect of prenatal testosterone. If sibling socialization is similar for twins and the CSS control group, then the effect of prenatal testosterone can be disentangled from the combined prenatal testosterone and socialization effect. Other independent variables included: age, age squared, family size, birth order dummies, maternal age at birth, nationality, test-year dummies, household type dummies, indicator of whether the mother was in DI in the year of giving birth, the mean test-score at the school the child attended in a given year, household income and a dummy for whether the mother was working. [82]

For boys with a male co-twin, no effects were found across all three measures – reading, maths, aggregate ability. For girls with a male co-twin, no effects were found for aggregate ability or reading but their Maths' scores were seen to be 7% of a standard deviation lower. As the authors point out, this finding is counterintuitive as it would be expected that girls with a twin brother, subject to increased levels of testosterone in utero, would demonstrate more typically male educational patterns with improved maths and worse reading performances. They also found no improvement in boys' math scores nor any worsening of their language scores. Given this, they conclude that they have found no evidence that prenatal testosterone amplifies (average) gender-specific

[81] This is to try to ensure similar, shared environmental experience for the control group.
[82] Both income and the working mother dummy related to the year in which the child turned four. Models with income controls were separately specified because of the effect on (reduction of) sample size.

differences. Further quantile regression analysis indicated that males subject to higher levels of prenatal testosterone were almost 2% less likely to be in the bottom 10% for both the aggregate ability and math scores. Girls subject to higher prenatal testosterone, however, were 2.5% more likely to be in the bottom 10% of the maths' test-score distribution. The negative impact of higher prenatal testosterone exposure on girls' Maths performance, the authors suggest, may result in the main from girls being proportionately more likely to be in the bottom 10% of the math-score distribution. Two potential explanations are offered for this counterintuitive finding. Firstly, that the relationship between performance in Maths and prenatal testosterone is a negative (rather than a positive) one and that boys overcome this disadvantage [though socialisation] by, for instance, playing with different toys. It is argued that a second alternative explanation might be that girls with a male co-twin are more masculine, and that this masculinity exerts a negative influence on educational performance at age twelve.

3.3.4 Heritability and Determinism

Krapohl et al. (2014) and de Zeeuwa et al. (2016) emphasize that, although they find variations in children's educational attainment to be the result of, in the main, innate differences, heritability does not mean determinism. They suggest that heritability in educational achievement can be mediated in the school environment through policy interventions. The more individually tailored approach to children's education, currently popular with policy makers (e.g. Curriculum for Excellence in Scotland) is favoured, as the interaction of genes with the school environment means that individuals will respond differently to various subjects and/or teaching methods used. Educational attainment can still be raised across the spectrum with uncomplicated policy interventions; de Zeeuwa et al. (2016) suggest that simply increasing teaching time in certain subject areas will raise average ability in such.

Genetic and environmental effects and their interaction, in addition to influencing overall educational and subject specific attainment, are likely to influence subject choice within the educational system; that is, there may be secondary as well as primary impacts. Focusing on students at the top of the ability distribution, De Philippis (2016), examines the effect of increased science provision in English secondary schools on the enrolment for and completion of STEM degrees. To increase enrolment in STEM subjects at university, the UK government introduced an entitlement to study advanced science at age 14 for high ability pupils in England and Wales in 2004. By 2011, the proportion of schools offering advanced science had risen to 80% from 20% in 2002, with the share of students opting to take this rising from 4% to 20% in the same period. She uses the timing of this reform and its unexpected implementation at school level (i.e. after students have made their secondary school selection) to address endogeneity issues. Three data sets are combined for the analysis; demographic characteristics from The Pupil Level Annual School Census (PLASC) are linked to Key Stage [83] attainment data from the NPD which in turn is linked to the Higher Education Statistical Agency (HESA) dataset to obtain degree/institution enrolment and completion information.

The results suggest that the probability of choosing to study science at age 16 increases by 5% if advanced science has been taken at age 14. Additionally, offering more science in secondary schools was associated with increased attendance at more prestigious universities and found to increase the probability of both STEM degree enrolment (1.5%) and completion (3%). However, the latter findings were driven entirely by increased post-school STEM participation by boys despite girls being as likely as boys to opt for advanced science in secondary school. As a result, the unintended consequence of this policy has been to widen the general STEM degree enrolment gender gap with girls opting to stick to female dominated subjects (such as Medicine). These gender differences in degree subject choice, she speculates, are likely to be correlated with different gender job/occupation preferences as found in other recent literature.

3.3.5 Summary

Behavioural Genetics' research, using twin studies, has been seen to demonstrate with a high degree of consistency that educational achievement is moderately to highly heritable. Innate individual traits combined with family influences have been seen to account for between 50% to 70% plus of the variation in educational achievement. Gender differences in numeracy and literacy related subject scores along traditional lines are evident in various contemporary studies of attainment, with boys tending to perform

[83] There are five Key Stages in the English Educational system:

better in the former and girls the latter (e.g. Machin & Pekkarinen, 2008). However, Behavioural Genetics' research suggests that there is no gender difference in general cognitive ability which is responsible, in the main, for determining educational attainment. This would indicate that differences in gender attainment across these broad subject domains are environmental and/or the result of gene-environment interactions. Twin studies in Behavioural Genetics have tended to focus on reading and, latterly, Maths with little investigation of the heritability of science (de Zeeuwa et al., 2016). Males and females make different STEM subject choices in secondary school (as noted above in Section 3.2). A largely unexplored issue is whether these choices are socially constructed (i.e. environmentally determined) or possibly genetic in nature with females and males having innate preferences for learning certain educational traits. This paper exploits the TTT hypothesis, following the work of Gielen & Zwiers (2018), but *uniquely* applies this to examine whether potential exposure to increased levels of testosterone might exert an influence on the in-school, STEM subject choices of females. The impact on STEM subject attainment is also examined. [84]

3.4 Data and Methodology

The SQA administrative data were rendered to identify twins and closely-spaced siblings (CSS). Twins (and other multiples) were identified principally through their common household id and date of birth. These variables were grouped in STATA [85] and *duplicates* identified. The derived household identifier variable used (*hid1*) was based on the strongest possible definition, created by grouping individuals' surnames, postcodes and first line of address. [86] The data were exported to Excel to create a control group of CSS based on common household and dates of birth within one year (365 days) of each other (as suggested by Tapp et al., 2011). Excel was used to identify the CSS group as it handles the calibration of dates far more easily than STATA. Additionally, its use enabled easy identification of twins born on consecutive days, increasing the sample of twins by

[84] As noted in the Introduction, only three other papers exploit TTT to investigate economic outcomes: Gielen & Zwiers, (2018) the only paper to date applying TTT to examine educational attainment, Gielen et al. (2016) investigating the gender wage gap and Cronqvist et al. (2015) examining gender differences in financial decision making.

[85] Using STATA's *egen* command.

[86] Two more weakly defined household identifiers were created also: hid grouping surname and postcode, hid2 grouping candidate postcode and first line of address.

another 162 individuals (81 further twin pairs), whilst at the same time purifying the control group (Table 3.3 below); something that has not obviously been carried out in other studies. Once the different birth date twins and CSS were identified and merged into the main data file, a separate Twins/CSS data file was created for the bulk of the analysis to speed up calibration time. The twins and CSS' data were refined further to identify the *mixed*- and *same-sex* pairings necessary for the analysis. As there is no information on zygosity as such, it was not possible to identify monozygotic (identical) twins, however, as the phenomenon of interest is the potential effect of TTT on subject choice, this is not particularly problematic. [87]

Table 3.3

Twin & CSS Sample Sizes

	Twins	**CSS**
Total Sample N	8,994	5,820
Male : Female (ratio)	4,379 : 4,615 (49:51)	2,917 : 2,903 (50:50)
Mixed : Same Sex (ratio)	2,442 : 6,552 (27:73)	2,236 : 3,584 (38:62)

Following Gielen and Zwiers (2016) and Evans and Martin (2000), higher order multiple birth observations were removed from the dataset since growing up as one of three/four identical/fraternal siblings is likely to be quite different from growing up as one of a pair. This reduced the twins' sample size by a total of 191 individuals; 61 sets of triplets and two sets of quadruplets. It should be noted that twins are more likely to come from more affluent two-parent households with higher earnings' capacity while CSS are more likely to come from relatively more deprived backgrounds (Gielen & Zwiers, 2018). Bhalotra and Clarke (2018), using data for the US and 68 developing countries, have shown that twins are more likely to be born to healthier, wealthier women bringing into question the validity of the twin instrument in instrumental variable studies. Examining the fertility–human capital trade-off, where twin births were used to instrument for fertility shocks, they found the twin-IV estimator to be biased upward as a result of this twin-birth sample selection. In the data used here, twins were significantly more likely to come from SIMD quintiles 4 and 5, the most affluent 40% of households, whilst CSS were significantly more

[87] Gielen & Zwiers (2018) suggest that the number of dizygotic twins can be approximated as twice the number of opposite-sex twins; this would imply a total of 4,884 dizygotic twins in the sub-sample, approximately 54%.

likely to come from SIMD quintiles 1 to 3, suggesting that sample selection may be a problem.

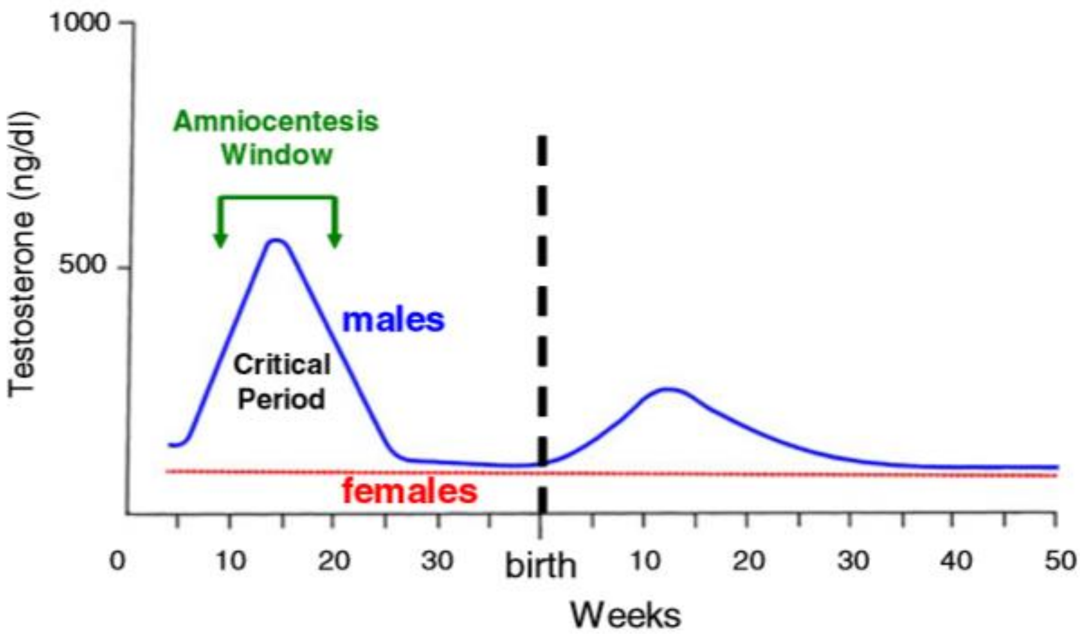
3.4.1 Methodology

The analysis is based on the premise that the TTT hypothesis holds and can be used to examine the gendered nature of STEM subject choices at all levels of secondary school qualification: SG/Intermediate 2, Higher, Advanced Higher. This is in keeping with Gielen & Zwiers (2018), who exploited TTT as an exogenous proxy to explore gender test score differences. Over half a century of animal studies has provided clear evidence of the effect of hormones on development whereby, for instance, masculinized neural development, cognition, and behaviour has been found in females treated with androgens, i.e. male sex hormones such as testosterone (Auyeung et al., 2013). It appears from animal research that the critical period for brain development in terms of sexual differentiation takes place when the differences in testosterone levels between the sexes are at their highest and it is believed that the same is likely to be true for humans. Between the eighth and twenty-fourth week of gestation male fetuses experience heightened levels of testosterone (see Figure 3.3), producing more than 2.5 times the levels found in females (*ibid*). After this, levels then fall with the difference between males and females becoming negligible until birth. Whilst males experience two further surges in testosterone - after birth (between 2 and 26 weeks, peaking around three-four months) and again in puberty, it is thought that the most significant effects on development are likely to occur within the in-utero window.

In addition to the development of the testes this in-utero exposure is critical to brain development (Tapp et al., 2011, Van de Beek et al., 2004 cited in Gielen & Zwiers, 2018). Prenatal testosterone would seem to determine the early organization of the brain, potentially influencing social and emotional behaviour, wiring it with masculine behavioural patterns in terms of preferences, personality, and temperament (Jordan-Young, 2010 cited in Gielen & Zwiers, 2018). This finding is supported by behavioural studies in humans, including those naturally exposed to higher testosterone levels, and generally by studies in nonhuman mammals (Auyeung et al., 2013). Testosterone levels in puberty, whilst also a predictor of behaviour, are believed to activate or fine-tune the earlier organisation of the brain and it is thought that this activation may be dependent on the initial, in-utero exposure when key tissues are first formed (*ibid*). Exposure to prenatal testosterone is much lower in females and levels tend to stay constant over the life-cycle (Tapp et al., 2011, Auyeung et al., 2013).

Figure 3.3

Circulating levels of testosterone in the human fetus and neonate



(Source: Auyeung et al., 2013, p558)

The TTT hypothesis suggests that for mixed-sex twins, significant amounts of testosterone can transfer from a male to a female co-twin as has been found in other litter bearing mammals. Many studies have shown mixed-sex twin pair females to exhibit more masculine physical, behavioural and cognitive traits. For instance, males are better in general at mentally rotating/visualising shapes and Heil et al. (2011) found females with male co-twins to be better at this than those with a same sex twin. Testosterone levels may be linked to brain volume [88] and right-handedness. Peper et al. (2009) found children who had a male co-twin (compared to those with a female co-twin) to have both larger total brain (+2.5%) and cerebellum (+5.5%) volumes with the largest volumes found in same-sex twin males followed by mixed-sex twin males, mixed-sex twin females, same-

[88] The brain volume of boys is approximately 10% larger than that of girls (Peper et al, 2010).

sex twin females. Same-sex twin males, theoretically, would be subjected to the highest prenatal levels of testosterone. Females with male co-twins have been found to exhibit significantly less left-handedness (Vuoksimaa et al., 2010). Tapp et al. (2011) and Peper et al. (2009) found no increased masculine behaviour or characteristics for males with a male co-twin. [89]

3.4.2 Causal Effect Identification

In keeping with the Potential Outcomes Framework (Angrist & Pischke, 2008, Rubin, 1974, Imbens & Wooldridge, 2009), the exposure of mixed-sex twin females to heightened levels of testosterone in utero may be regarded as the application of an exogenous treatment. There are two outcomes of interest: STEM subject choice and STEM subject attainment. The primary aim is to ascertain whether females exposed to increased levels of testosterone might make more "masculine" (STEM) subject choices and, therein, subsequently exhibit different levels of attainment in these subjects compared to other "untreated" females. Therefore, in terms of potential outcomes, STEM subject choice (for instance) can be expressed as follows:

$$Y_i = y_{1i}\, if\, d_i = 1, y_{0i} if\, d_i = 0$$

$$= y_{0i} + (y_{1i} - y_{0i})d_i$$

Where:

Y_i is the measured outcome of interest for individual i, is the probability of choosing a given STEM subject, that is $Y_i = 1$

y_{0i} is the baseline potential outcome probability of choosing a STEM subject

y_{1i} is the potential treatment outcome probability of choosing a STEM subject

$d_{i=1}$ indicates the increased in-utero exposure to testosterone for female twins with male co-twins, that is the *treatment*

$d_{i=0}$ indicates *normal* levels of testosterone exposure for females in-utero, ie not treated

The change in the probability of selecting a STEM subject (y_{1i}- y_{0i}) is the causal effect of interest but this is unobservable as it is impossible to observe the two outcomes for a

[89] See Gielen & Zwiers (2018) for other studies.

single individual. The *second best* option is to compare outcomes for different individuals; that is, those who have been *treated* against a control group of *like* but, otherwise, *untreated* individuals. This allows the observed effect, in terms of the observed difference between the means for the treated and untreated groups to be estimated, viz:

$$\Delta Y_i | d_i = E[Y_i | d_i = 1] - E[Y_i | d_i = 0]$$

If non-random treatment selection exists, however, this may produce biased results that either lead to the incorrect conclusion that a treatment was (in)effective or to the significant under-/overestimation of the true treatment effects depending on the direction of the bias (Bogard, 2013). Non-random selection can occur if the potential baseline outcomes (Y_{0i}) of the treated and control groups are not the same and can be expressed as follows:

$$\Delta Y_i | d_i = 1 = E[Y_i | d_i = 1] - E[Y_i | d_i = 0] = \underbrace{E[Y_{1i} - Y_{0i}]}_{ATE} + \underbrace{E[Y_{0i} | d_i = 1] - E[Y_{0i} | d_i = 0]}_{SB}$$

Here the observed effect or difference is equal to the population average treatment effect (ATE) plus a term arising from selection bias (SB). If the potential baseline outcomes (Y_{0i}) of the treated and control groups are not significantly different, then the selection bias term would tend to zero, and the observed difference would represent the population average treatment effect. However, if the potential baseline outcomes (Y_{0i}) differ significantly for those who are treated (d_i=1) from those who are not treated (d_i=0), then the selection bias term may take a positive or negative value, producing a biased result (*ibid*). If this is the case, when the observed difference between treated and untreated groups is estimated, selection bias becomes confounded with the ATE. In turn, if the SB term is sufficiently large, it can dominate the actual treatment effect, resulting in the conclusion that the treatment was (in)effective or either the under-/overestimation of the true treatment effects (*ibid*).

As noted above, the twins in the data were found to be significantly more likely to come from SIMD quintiles 4 and 5, the most affluent 40% of households, whilst the CSS group was significantly more likely to come from SIMD quintiles 1 to 3 (Table 3.4). Individuals from SIMD households 4 and 5, therefore, are more likely to have been "treated" than those from SIMD 1 to 3 households. This could confound the effect of increased exposure to testosterone with socio-economic background effects, potentially resulting in an overestimation of any positive effect of testosterone on subject choice. For instance, as twins are more prevalent among SIMDs 4 and 5, they are more likely to have greater levels of STEM social capital in terms of family/family friends working in STEM or related occupations which is known to influence STEM subject uptake in schools (Archer et al., 2013). It was also found that there were significant differences in the geographical distribution of twins across LA areas. Given the nature of the administrative data used, it is not possible to meaningfully instrument for either SIMD or LAs. Therefore, the issue of endogeneity arising from any selection bias in the analysis has to be left moot.

Table 3.4

Twins & CSS by SIMD Quintile

SIMD	1	2	3	4	5	Total
Twins: No	1,590	1,646	1,696	1,922	2,022	8,876
%	56.34	45.82	59.47	73.22	75.48	60.92
CSS: No	1,232	1,946	1,156	703	657	5,694
%	43.66	54.18	40.53	26.78	24.52	39.08
Total	2,822	3,592	2,852	2,625	2,679	14,570

Pearson χ^2 (4) = 776.3899 Pr = 0.000

Given TTT and following Gielen & Zwiers (2018), four further assumptions must be made. Firstly, that sex is distributed randomly among twins. Secondly, more strictly, that a co-twin's sex is only related to its twin sibling's educational outcomes through TTT. Thirdly, that socialization is the same for identical and fraternal twins. This is necessary as zygosity cannot be determined and, therefore, no direct distinction can be made between genetic and environmental reasons for sibling similarity (Leckie, 2010). Finally, twins must be representative of the general population.

Given the distributions of twins across the SIMD quintiles in favour of 4 and 5, an over-representation of mixed-sex twins in SIMDs 1-3 might have a compensating effect of spreading the "treatment" more evenly across SIMDs. On the other hand, an over-representation of mixed-sex twins in SIMDs 4 and 5 would have the effect of further concentrating the treatment. In the event, it was found that the distribution of mixed-sex twins within the SIMD quintiles and, therefore, importantly, female co-twins exposed to heightened levels of testosterone was not significant (Table 3.5). Ergo, the distribution of sex amongst twins across SIMD quintiles is random but, as discussed, the treatment is not to the extent that there is an over-representation of twins amongst the most affluent 40% of households.

Table 3.5

Mixed- & Same-sex Twins by SIMD Quintile

SIMD	1	2	3	4	5	**Total**
Same-sex Twins: No	1,190	1,188	1,230	1,390	1,462	6,460
%	74.84	72.17	72.52	72.32	72.30	72.78
Mixed-sex Twins: No	400	458	466	532	560	2,416
%	25.16	27.83	27.48	27.68	27.70	27.22
Total	1,590	1,646	1,696	1,922	2,022	8,876

Pearson χ^2 (4) = 4.2110 Pr = 0.378

As Gielen & Zwiers (2018) point out, it is almost certain that the second assumption does not hold; any comparison of the educational outcomes of twins with male co-twins with those of twins with female co-twins measures *both* the effect of any prenatal testosterone exposure and the effect of growing up with an opposite- or same-sex sibling. Truncating the data to include only twins means that it is necessary to consider whether or not such a sample might be representative of the general population. It has been asserted that twins may have lower intellectual skills and educational achievements than singletons and that this may be a result of a link between lower birth weight and intellect/educational performance (see, for instance, reviews of studies by Deary, 2006 and de Zeeuwa et al., 2016). In addition, it has been suggested that intrauterine growth retardation and shorter gestational length may result in twins having lower intellectual skills than singletons (Cohen et al., 2002). If so, this might also confound any testosterone effect, potentially resulting in underestimation if the effect were positive.

The evidence that twins have lower intellectual skills is evaluated by Deary (2006) who reviews a number of studies. Studies of Scottish twins, using historic data sets, have indicated that twins may have lower intellectual abilities (*ibid*). The IQs of 11-year olds born in 1921 and 1936 were reported, respectively, in the Scottish Mental Health Surveys of 1932 and 1947. For both surveys, twins were found to score approximately 5 IQ points lower than single children (one third of a standard deviation) when controlling for father's social class, overcrowding, childhood height, school attendance and family size (Deary et al., 2005 cited in Deary, 2006). A study focusing on children in Aberdeen in the 1950s, reported a similar difference in the IQ test scores of 7- and 9-year old twins and singletons (Ronalds et al., 2005 cited in Deary, 2006). Danish and Dutch studies (Christensen et al., 2006 and Posthuma et al., 2000 both cited in Deary, 2006), respectively testing twins at age 16 and in adulthood, however, found no significant differences between twins and singletons. The evidence, Deary asserts, suggests that while differences in ability and educational performance between twins and singletons may exist as late as age 11, they have dissipated by 16 (*ibid*). This is further supported by the review of several studies by de Zeeuwa et al. (2016,), suggesting that differences in twins' and singletons' body size, general cognitive ability and educational achievement dissipate early in childhood. With respect to the issue of a birth weight-IQ relationship, Deary argues that this would apply also to singletons. Whilst twins may be lighter on average, they will not necessarily differ from singletons in ability in the same way that men and women differ little in mean intelligence despite there being some evidence of a link between brain size and IQ: men having, on average, a larger brain mass. A Dutch study (Cohen et al., 2002) of the educational achievement of female twins found no significant difference between the performance of the former and the Dutch female population as a whole. These reviews of such evidence suggests that it is reasonable to assume that the sub-sample of twins is representative of the population in terms of intellectual ability by school certification age.

Nevertheless, the CSS control group is used to mitigate for the likely violation of assumption two (the sex of a co-twin is only related to its twin sibling's educational outcomes through TTT) and the possibility that twins may not be representative of the general population. This should allow for the comparison of the potential STEM subject choice impact (if any) of heightened testosterone exposure for mixed-sex twin females within a general sample population of similarly socialised children.

Sibling similarity in academic achievement and IQ has been demonstrated in family studies (Leckie et al., 2010), however, socio-economic outcomes can vary considerably within families if, for instance, children are brought up at different times or parental resources are allocated unequally between them, (Nicoletti & Rabe, 2010) and there is some evidence to suggest that educational differences across families are more likely to be correlated with differences in ability than is the case for twins (Bonjour et al., 2002). The twin method assumes that twins brought up together are similar as a result of both the effects of shared genes and shared environmental factors (Krapohl et al., 2014).

The CSS group is incorporated in the analysis on the assumption that their socialization is similar to that of twins; being born within a year of each other should mean that, like twins, they experience similar environments. This is the Equal Environments Assumption (EEA) which is applied here to assume that there are no systematic differences in the upbringing environments of identical or fraternal twins and CSS that influence school educational outcomes. If the EEA holds, then differences between twins would be determined by differences in their genetic makeup; identical twins share 100% of their genes while fraternal twins share approximately 50% as do non-twin siblings (*ibid*). The EEA could be invalid if, for instance, MZ twins are treated more similarly than DZ twins and, indeed, CSS. This would mean than any differences in similarity between twin pairs would no longer be the result of genes alone but a combination of genes and environmental upbringing which would lead to upward bias on any gene effect (de Zeeuwa et al., 2016). Previous research suggests that the EEA holds for both general cognitive ability and educational achievement (Evans & Martin, 2000, de Zeeuwa et al., 2015, de Zeeuwa et al., 2016).

The EEA must be considered also in the context of the school environment. The relative importance of the school environment and genes in explaining variations in educational achievement is difficult to untangle if both vary. The existence of a national curriculum, as was the case in Scotland at SGI level, restricts variation in school environments, ceteris paribus, and would suggest an increased relative contribution of genes to the variation in educational achievement (Kovas et al., 2013). As noted in Chapter Two, the quality of provision varies little from school to school in Scotland (OECD, 2007, Commission on School Reform, 2013), suggesting that other factors influence attainment and educational choices.

3.4.3 Model Specifications

The aim was to explore whether the STEM subject choices and attainment of females from mixed-sex twin pairs, potentially exposed to heightened levels of testosterone in utero, might be different from those of other females. STEM subject choice was examined at all three levels of qualification - SGI, H and AH, firstly for the population data and the Twin/CSS sub-sample for comparison purposes and then for Twin/CSS females only to ascertain the presence of any testosterone effect. As STEM attainment for the population was previously examined in the Chapter Two investigation into attainment in facilitating subjects, it is examined here, at all three levels of qualification, just for the Twin/CSS sub-sample and then again for Twin/CSS females only to determine the presence of any testosterone effect. The analysis employed the same variables used in the Chapter Two analysis viz:

- Sex (Male=1)
- SIMD 1–5: individuals' household Scottish Index of Multiple Deprivation 2009 quintile
- SGI UCAS points: total Standard Grade/Intermediate 2 UCAS points at age-16 (Highers' models)
- Higher UCAS points: total Highers' UCAS points at age-17 (Advanced Highers' models)
- School quartiles 1-4: individuals' SGI UCAS points' school quartile (Highers' and Advanced Highers' models)
- SGI/Higher/AH cohort size: school size as measured by Standard Grade/Intermediate 2, Higher or Advanced Higher cohort size respectively
- % pupils SIMD 1 & 2: schools' socio-economic composition given by the percentage of pupils from SIMD quintiles 1 and 2, the most deprived 40% of households
- Urban/Rural: dummy variables for urban/rural location and degree of remoteness
- LA: local authority dummy variable for individuals' school
- Year: dummy variables for years 2002-2009

In addition, the following Twin/CSS dummy variables replaced Sex in the Twin/CSS female-only subject choice and attainment model specifications:

- twin_msf – female from a mixed-sex twin pair
- css_msf – female from a mixed-sex, closely spaced sibling pair

STEM subject choice was modelled using binary logistic regression as outlined in Chapter One. Individual dependent variables taking the form 0/1 (no uptake/uptake) were created for each STEM subject, at each qualification level. Accordingly, separate logistic regressions were estimated for each individual STEM subject at each qualification level (16 separate models) for the entire population dataset, the Twins/CSS sub-sample data and Twins/CSS females only. [90] Sex was replaced in the Twins/CSS female-only models by the two dummy variables above, taking the generic form:

$$ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 twin_m sf + \beta_2 css_m sf + \sum_{j=3}^{7} \beta_j SIMD1 - 5_i + \beta_8 SGI\ cohort\ size_i + \beta_9 \%\ pupils\ SIMD\ 1\ \&\ 2_i + \sum_{k=10}^{15} \beta_k UrbanRural_i + \sum_{l=16}^{47} \beta_l LA + \sum_{m=48}^{55} \beta_m Year + \varepsilon_i$$

This is the likelihood that individual *i* will choose a given STEM subject at SGI level. Similar to Chapter Two, the Higher and Advanced Higher STEM subject choice models also included measures of individuals' previous absolute and relative (within school) attainment, that is: total SGI UCAS points for the Highers' models, total Higher UCAS points for the Advanced Highers' models, SGI UCAS points-based School quartile for both models.

The full Twins/CSS sub-sample data models were estimated to the provide benchmark results and to ascertain if the sub-sample was representative of the general population. The female-only Twins/CCS data models were designed to explore specifically whether females exposed to heightened testosterone levels might make different STEM subject choices from other females. Mixed-sex CSS pair females then provided the control group since they shared the same percentage of common genes with their brothers (approximately 50%) and the same socialisation experience as mixed-sex, twin pair females but, crucially, not the exposure to increased testosterone levels.

In keeping with the analysis of attainment in Chapter Two, multinomial logit regression analysis was used to explore STEM subject attainment specifically for the whole

[90] In all, a total of 48 models were estimated; 16 for each cut of the data.

Twins/CSS' sub-sample, for comparison with the population data, and for Twin/CSS females to attempt to identify potential genes/socialisation effects. Multinomial logits were estimated for each of the STEM subjects at all levels of qualification (SGI, H, AH) with the individual dependent variables having the same four-outcome structure employed in the earlier attainment analysis (outcome 1: fail, outcome 2: low pass, outcome 3: middle pass, outcome 4: high pass). As with the specification of the subject choice models, the multinomial logit attainment models were estimated firstly using the whole Twins/CSS sub-sample with the generic model given by:

$$ln\left(\frac{p_{i2}}{p_{i1}}\right) = a_2 + \beta_{2|1}Sex_i + \sum_{j=2}^{6}\beta_{2|j}SIMD\ 1-5_i + \beta_{2|7}SGI\ cohort\ size_i$$
$$+ \beta_{2|8}\%\ puplis\ SIMD\ 1\ \&\ 2_i$$
$$+ \sum_{k=9}^{14}\beta_{2|k}UrbanRural_i + \sum_{l=15}^{46}\beta_{2|l}LA + \sum_{m=47}^{54}\beta_{2|m}Year + \varepsilon_i$$

This is the likelihood that individual *i* will obtain a low pass (p_{i2}) as opposed to a fail (p_{i1}) in a STEM subject at SGI level. The Higher and Advanced Higher STEM attainment models also included the measures of individuals' previous absolute and relative attainment (SGI UCAS points for Highers' models, Higher UCAS points for Advanced Highers' models and SGI UCAS points-based School quartile for both models). [91] Again, in the Twin/CSS female-only models, sex was replaced by the two mixed-sex twin/CSS female dummy variables included, to ascertain whether females exposed to increased testosterone levels might perform differently in STEM subjects from other females.

Summary statistics for the data used in the logistic subject choice and MLM attainment models are provided in Appendix C, Tables C1 and C2 respectively. The working samples for the population subject choice models varied from the total number of individuals observed in the data (489,468) at age 16 as a result of incomplete/inconsistent data entry (as noted in the Introduction, Administrative Data) and, subsequently, were reduced further naturally as the numbers of individuals proceeding to senior school level qualifications falls. The working samples for the Twins/CSS subject choice and attainment

[91] In the event, the inclusion of LA fixed effects and, at Higher, individuals' school quartile caused convergence issues whereby only a small number of observations were completely determined. Accordingly, these were removed from all the Twins/CSS attainment models.

models, *de facto*, were limited to the identified numbers of Twins and CSS as reported above in Table 3.3. Again, working sample numbers were reduced initially from the total number of identified Twins/CSS for the subject choice models as a result of incomplete/inconsistent data entry and then fell further with the reduced uptake of senior school level qualifications. The Twins/CSS attainment models' sample numbers varied according to whether or not individuals had been entered for the subject in question at the particular level. As in both the earlier chapters, the working samples for all models were comprised of those individuals who had taken the specific qualifications *on-time*; that is SGI qualifications in S4, Highers in S5 and Advanced Highers in S6. School level, cluster-robust standard errors are employed for all models to account for the possibility that model errors for individuals in the same school might be correlated.

3.4.4 Predictive Power of Models

Appendix C, Table C3 shows the results of the classification and goodness of fit tests for the logistic regression subject choice models. At SGI level, across all STEM subjects and the three datasets (whole population data, Twins/CSS sub-sample, Twins/CSS female-only), in general, 70% or more observations were correctly classified. This rose to 90% for general Science (all datasets) but was as low as 63% for the Biology, female only model. Calculation of the Adjusted Count R^2 generally demonstrated a reduction in the classification error rate compared to predictions based only on the dependent variable's marginal distribution but these were sometimes negligible. All models passed the goodness-of-fit test based on the Pearson chi-squared statistic is significant at the 0.001 level.[92]

At Higher level, across all subjects and datasets, 70% or more observations were correctly classified. The widest range occurred among the female-only Higher level models with 70% of observations correctly classified for the Biology model and 93% for Computing. Calculation of the Adjusted Count R^2 revealed a wide range in the reduction of classification error rates, from 0% for Computing (population data model) to 117% for

[92] The Pearson chi-squared statistic evaluates the sum of the square of the Pearson residuals across all covariate patterns; it tests the hypothesis of conformity between the predicted and observed frequencies across the covariate patterns (Kohler & Krueter, 2009) The Pearson residuals compare the number of successful classifications of individuals having particular covariate patterns with the predicted number of successes having those covariate patterns.

Biology (also population data model). The zero or negligible reductions in classification error rates occurred where positive observations were relatively sparse, for example Computing in the population data, Computing and Physics in the Twins/CSS female-only models. For the Twins/CSS data, all versions of the Higher Computing models and the female-only Biology and Physics models failed the goodness of fit test. The AH subject choice models could not be estimated with the female-only Twins/CSS sub-sample. The AH models for the population and the whole Twins/CSS sub-sample exhibited no reduction in classification error rates compared to predictions based only on the dependent variables' marginal distributions and generally failed the goodness of fit tests.

As with the facilitating subject attainment models in Chapter Two, the Twins/CSS STEM subject attainment models were estimated originally as ordered logits but, in general, failed the proportional odds' test (Appendix C, Table C4). In the few cases where the proportional odds' test statistics were insignificant at the 5% level (indicating that ordered logit models could be used), they were significant at the 10% level. To enable comparison across all STEM subjects at the different levels, the STEM attainment models were re-estimated as multinomial logits and augmented by average marginal effects (AMEs).

3.5 Results and Analysis

The uptake of STEM subjects by gender at the different qualification levels is shown in Figures 3.4 to 3.6 for the entire dataset and Figures 3.7 to 3.9 for the Twins/CSS' sub-sample. Engineering is not included as it is rarely offered at secondary school level; between 2000 and 2009, there were just 27 entries for Intermediate 1 Engineering and 86 for Intermediate 2. When examining Figures 3.4 and 3.7, it should be remembered that the structure of the SG-based curriculum (before the introduction of CfE) was based on closed option choices such that, in addition to Maths, the study of a named or general Science and a technology related subject was compulsory for the vast majority of fourth year pupils. Nevertheless, it can be seen from Figures 3.4 to 3.6 that the uptake of particular STEM subjects is highly gendered at all levels. Biology is the "female" science of choice with uptake essentially in the ratio 30:70, male to female, at all levels. Physics is the "male" science of choice and is the mirror image of Biology at 70:30, male to female, at SGI and H levels but becoming even further male dominated at AH level (80:20, male to female). Maths is taken by equal numbers of male and female students until AH level where uptake shifts towards males (61:39), while Chemistry is taken by equal numbers at all levels. Computing is also male-dominated and becomes increasingly so as qualification level rises; 64:36 at SGI level rising to 85:15, male to female, at AH level. Figures 3.7 to 3.9 show that the same pattern of STEM subject uptake is replicated in the Twins/CSS' sub-sample. The SQA statistics for 2010-2016 (Scottish Government, 2017c) suggest that this gendered pattern of STEM subject choice has not changed.

Figure 3.4

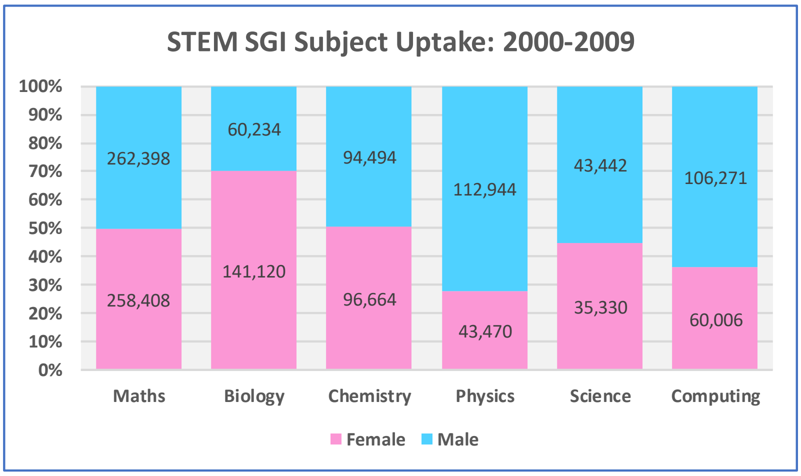


Figure 3.5

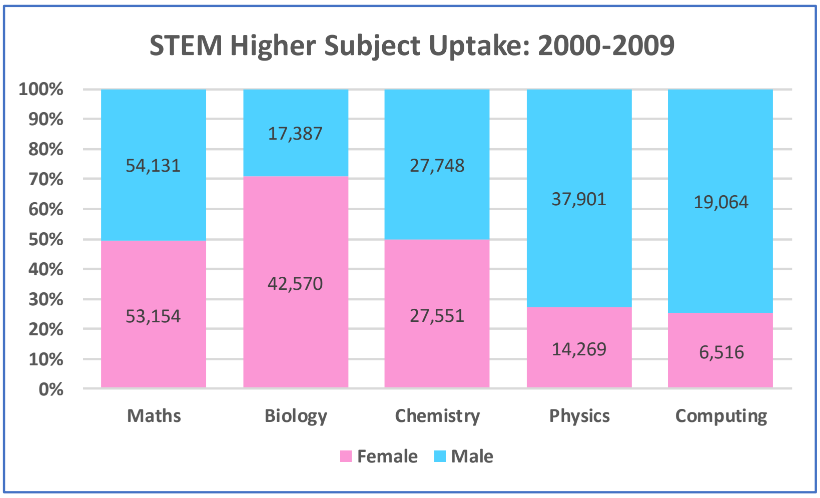


Figure 3.6

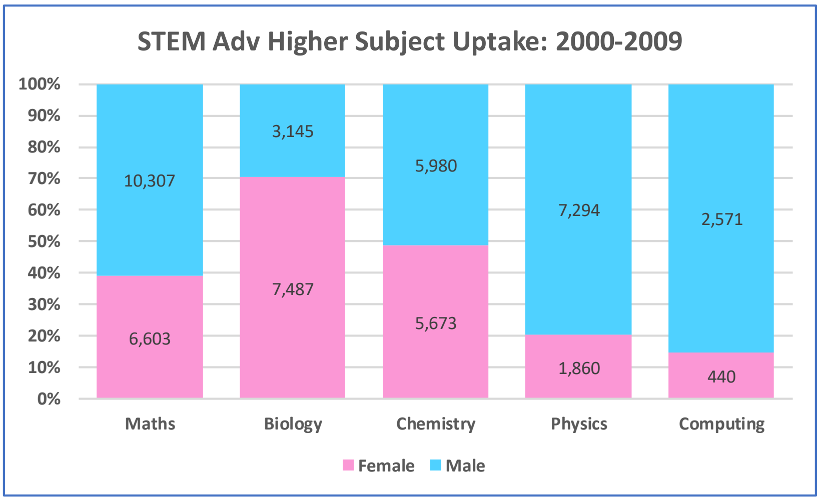


Figure 3.7

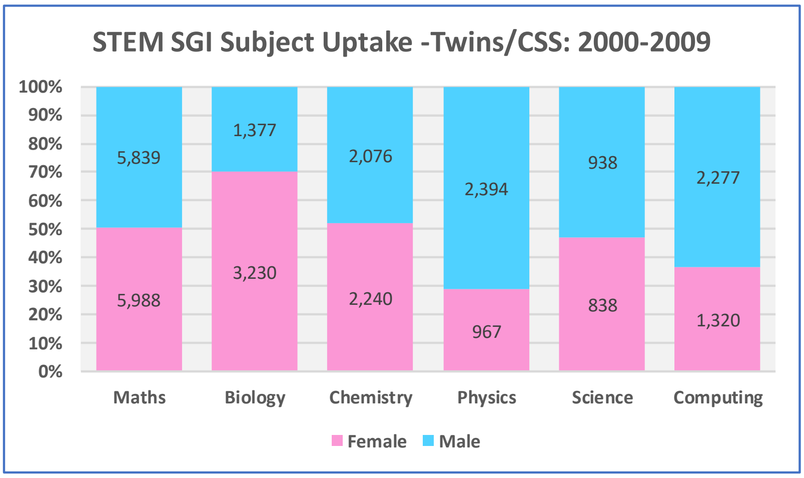


Figure 3.8

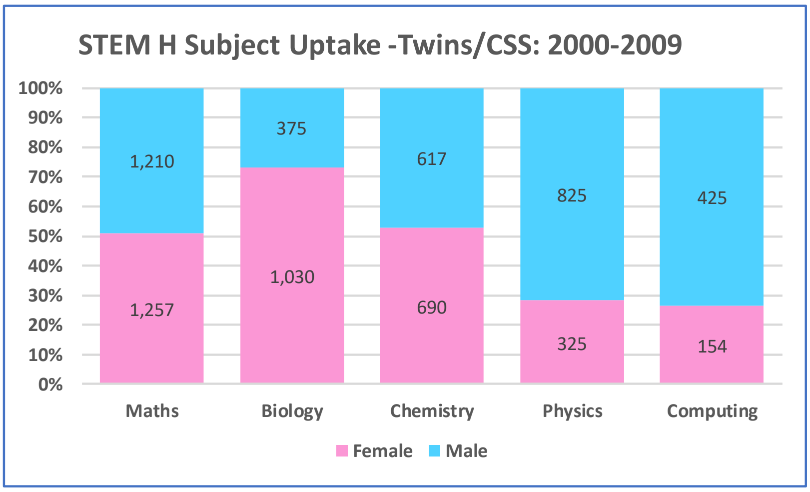
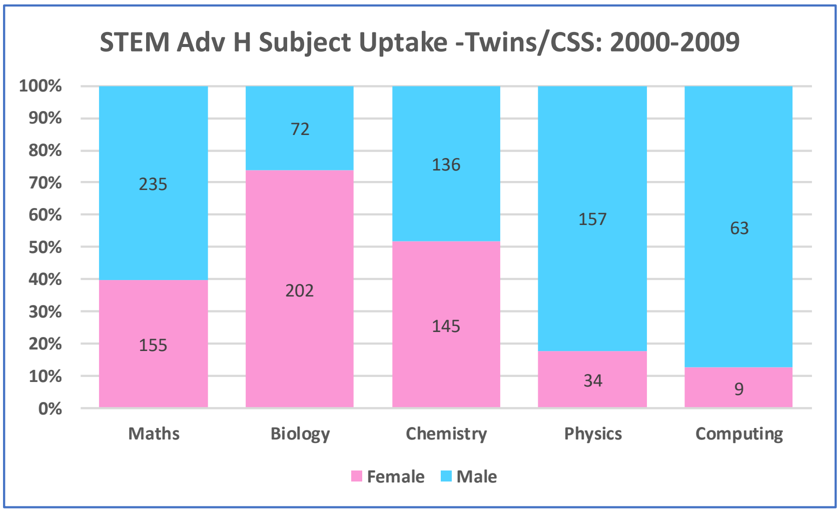


Figure 3.9



3.5.1 STEM Subject Choice – All Qualification Levels: Population Data

To check that the Twins/CSS sub-sample data were representative of the population, the STEM choice models were estimated for the entire dataset at SGI, Higher and Advanced Higher levels first. The results of these models are shown, respectively, in Tables 3.6, 3.7 and 3.8 for the main variables of interest. At SGI level (Table 3.6), it can be seen that Physics, Computing and general Science are more likely to be *male* choices. Males are more than three-and-a-half times as likely as females to opt for Physics as females (odds's ratio (OR): 3.6) and twice as likely to pick Computing (OR: 2.2). They are considerably less likely to take Biology [93] (OR: 0.27) and 2%-6% less likely to take Maths and Chemistry respectively. Boys are 25% more likely to opt for general Science that girls. It should be noted, however, that under the SG curriculum where students have to take at least one science, general Science tended to be a *fall-back* choice for academically weaker individuals or those who were not interested in the sciences. This result suggests that, in keeping with the earlier findings in Chapter One and the work of Machin and Pekkarinen (2008), aggregate male performance may be bimodal with high and low achievers.

With respect to individuals' socio-economic background, a distinct *mirror image* pattern is evident in the choice between a named science and general Science. The lower the SIMD quintile that an individual is from, the less likely they are to opt for a named science compared to someone from SIMD quintile 5, the least deprived (most affluent) 20% of households. On the other hand, the lower an individual's SIMD quintile, the more likely they are to have opted for general Science compared to someone from SIMD 5, ranging from 54% more likely (SIMD quintile 4) to approaching four times as likely (SIMD quintile 1, OR: 3.69). Individuals from lower SIMD quintiles are less likely also to have taken SGI level Maths and/or Computing, although, notably, the relative odds' ratios appeared to have been less detrimental for Computing. In general, schools' socio-economic composition is significant, but their SGI cohort size is not. A 1% increase in the proportion of a school's pupils from SIMDs 1 and 2 (the most deprived 40% of households), reduces the likelihood of an individual taking one of the STEM SGIs by 1%-2%.

[93] Whilst, in general, Maths was compulsory under the SG curriculum structure, very weak students would have been more likely to be presented for Access level qualifications below SGI level.

Table 3.6

STEM Subject Choice at Standard Grade / Intermediate 2 : Population

	Maths	**Biology**	**Chemistry**	**Physics**	**Science**	**Computing**
Sex (M=1)	0.976	0.274***	0.937***	3.629***	1.248***	2.184***
SIMD 1	0.558***	0.493***	0.350***	0.355***	3.690***	0.699***
SIMD 2	0.664***	0.630***	0.480***	0.484***	2.906***	0.838***
SIMD 3	0.701***	0.712***	0.587***	0.598***	2.161***	0.899***
SIMD 4	0.858**	0.856***	0.751***	0.764***	1.538***	0.905***
SGI cohort size	1.005***	1.001	1.001	1.001	1.000	1.001
% pupils SIMD 1 & 2	0.982***	0.992***	0.992***	0.992***	1.001	0.995**
Observations	444900	444900	444900	444900	442706	444900
Log likelihood	-184441.34	-253033.79	-259320.53	-222338.28	-125724.18	-238748.26
Wald chi2(50, Sci 49)	1137.53***	9663.88***	3946.99***	10697.03***	1784.53***	3315.25***
Pseudo R^2	0.1398	0.0922	0.0466	0.1008	0.1093	0.0612

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (356 schools).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As noted in Chapter One, the Scottish state secondary school system becomes *open (*as opposed to closed) option choice at Higher level although subject access does depend on previous SGI performance, teacher assessment of students' ability to succeed at Higher level and school/LA provision. Subject choice at Higher is of critical importance in determining both HE entry (vis a vis *facilitating* subjects) and degree course and, therefore, potential career options. Not choosing STEM subjects at Higher shuts off access to related degrees at university and subsequent, relatively well-paid careers (Scottish Government, 2017f). STEM subject choices at Higher for the whole dataset are presented in Table 3.7. Uptake of STEM subjects at Higher is very clearly gendered with males being more likely to take every subject apart from Biology; where they are only 39% as likely as females to study this in their fifth year at school. Males are 60% more likely to take Chemistry than females, twice as likely to take Maths, four times as likely to take Computing and almost seven times as likely to choose Physics. Previous SGI attainment matters with individuals being between 1% (Computing/Biology) and 4% (Maths) more likely to take a STEM Higher for a one-point increase in their total SGI UCAS' points.

The impact of socio-economic background is considerably reduced for STEM subject choice at Higher (Table 3.7). It is not significant at all for Biology. For Chemistry, it is only being from SIMD quintile 1 that is significant; compared to those from SIMD quintile 5, individuals from the most deprived 20% of households are 9% less likely to opt for Higher Chemistry. In Maths and Physics, whilst the usual pattern of impact is still evident, the effect is much reduced. Individuals from SIMD quintile 1, for instance, are 14% less likely to take Higher Maths and 9% less likely to take Higher Physics compared to someone from SIMD quintile 5; a greatly reduced impact from that seen at SGI level (in Table 3.6) where they were, respectively 44% and 64% less likely to take these subjects. This suggests that the impact of socio-economic background, at least on subject choice, *wanes* given previous levels of attainment. For Computing, individuals from SIMDs 2-3 were between 10% and 14% more likely to take the Higher than those from SIMD 5 (the least deprived 20% of households). Schools' socio-economic composition is significant only for Biology, indicating that the likelihood of choosing this subject is reduced slightly (less than half of one percent) by a one percent increase in the

Table 3.7

STEM Subject Choice at Higher : Population

	Maths	Biology	Chemistry	Physics	Computing
Sex (M=1)	2.181***	0.385***	1.599***	6.889***	4.406***
SGI UCAS points	1.038***	1.013***	1.025***	1.025***	1.006***
SIMD 1	0.861***	0.970	0.912**	0.914*	1.072
SIMD 2	0.885***	0.962	1.010	0.908***	1.100**
SIMD 3	0.908***	1.012	1.009	0.938*	1.135**
SIMD 4	0.959	1.006	0.975	0.952**	0.998
Higher cohort size	1.000	1.001	0.999	0.999	1.000
% pupils SIMD 1 & 2	0.998	0.996*	0.998	0.997	0.996
Observations	185698	185698	185698	185698	185698
Log likelihood	-83925.982	-91530.821	-80904.34	-72499.63	-54819.072
Wald chi2(54)	11625.25***	6621.27***	6465.24***	12355.68***	4511.18***
Pseudo R^2	0.3400	0.1153	0.1902	0.2454	0.1173

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Individuals' SGI performance school quartile, Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (348 schools).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

percentage of pupils coming from the most deprived 40% of households (SIMDs 1 and 2). The size of schools' Higher cohorts is insignificant across all subjects.

As might be expected, the gendered pattern of Higher STEM subject choice is repeated at AH level (Table 3.8); this is likely to be the case since an individual cannot take a given Advanced Higher unless they already have a pass at Higher in the subject in question. Compared to females, males are 25% more likely to take AH Chemistry, twice as likely to take AH Maths, almost six times as likely to take AH Physics and seven times as likely to take AH Computing while being only 42% as likely to take AH Biology. Individuals' total Higher UCAS points were significant for all subjects but, counter-intuitively, were seen to exert a small negative influence on uptake, at most a 0.5% (for Physics) decrease in the likelihood of taking a STEM AH for one extra Higher UCAS point. This might be because those that have achieved good Highers have secured enough UCAS points already for their chosen HE route and so do not need further points at AH level.

No socio-economic background effects were present in the uptake of Biology or, in general, Maths and Computing. AH Chemistry was the only subject choice for which there were significant SIMD effects at all levels with the familiar pattern of influence (i.e., the lower an individual's SIMD, the less likely they are to take a subject); an individual from SIMD quintile 1 was only 64% as likely as someone from SIMD 5 to choose AH Chemistry. Individuals from SIMD quintiles 1 and 2 were significantly less likely to choose AH Physics (75% and 89% respectively). The size of school's AH cohorts was significant only for Biology, exerting a small positive impact on the uptake of Biology. A school's percentage of pupils from the most deprived 40% of households (SIMDs 1 and 2) was only significant in the uptake of AH Biology and Chemistry with a small negative impact on both subjects. For a one percent increase in a school's percentage of pupils from SIMDs 1 and 2, the likelihood of choosing AH Biology or Chemistry fell by approximately half a percent.

Table 3.8

STEM Subject Choice at Advanced Higher : Population

	Maths	Biology	Chemistry	Physics	Computing
Sex (M=1)	2.138***	0.422***	1.253***	5.892***	7.295***
Higher UCAS points	0.998***	0.998***	0.996***	0.995***	0.997***
SIMD 1	0.925	0.958	0.635***	0.747***	1.175
SIMD 2	0.877**	0.981	0.822***	0.886*	1.183*
SIMD 3	0.965	1.056	0.911*	0.933	1.003
SIMD 4	1.004	0.990	0.909*	0.976	0.931
AH cohort size	0.999	1.005***	0.998	1.003	1.003
% pupils SIMD 1 & 2	1.003	0.996*	0.994**	0.997	1.002
Observations	53676	53676	53676	53676	53200
Log likelihood	-25521.415	-20287.465	-20153.781	-15268.757	-7688.1341
Wald chi2(54, Comp 52)	2328.36***	2786.49***	1623.64***	2756.38***	13173.60***
Pseudo R^2	0.0632	0.0515	0.0479	0.1248	0.1510

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Individuals' SGI performance school quartile, Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (340 schools reducing to 336 for Computing).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5.2 STEM Subject Choice – All Qualification Levels: Twins/CSS Data

Logistic regression results for STEM subject choice at SGI, Higher and Advanced Higher levels are shown for the whole Twins/CSS sub-sample in Tables 3.9, 3.10 and 3.11 respectively. The main variables of interest are shown in these tables. The highly gendered nature of Biology, Physics and Computing, evident in the basic descriptive statistics for subject uptake and population data models, is reflected in the Twins/CSS sub-sample regression results. There was no significant gender effect for Maths or Chemistry. Males are only 28% as likely to take Biology at SGI level compared to females but over three-and-a-half times as likely to study Physics and twice as likely to opt for Computing. There were strong, significant socio-economic background effects for Maths and the three named sciences. The likelihood of taking a named science reduced sequentially as SIMD quintile fell; those from the most deprived 20% of households were only between 36% (Chemistry) and 45% (Biology) as likely to take these compared to an individual from the most affluent, top SIMD quintile 5. This is mirrored in the uptake of general Science with those in the bottom SIMD quintile almost 4.5 times as likely to take this as those from the most affluent 20% of households. As above, under the SG-based curriculum, pupils were required to take at least one science, either named or general. The socio-economic composition of the school is significant, indicating that for every 1% increase in the percentage of pupils from the most deprived 40% of households, the likelihood of taking a named science at SGI level falls by 1% (2% for Maths).

At Higher, (Table 3.10), the only variables that were significant across all STEM subjects are gender and SGI points. Males were significantly more likely to take all Higher STEM subjects apart from Biology. Compared to females, they were over seven times more likely to take Physics, twice as likely to take Maths and 36% more likely to take Chemistry but only 33% as likely to take Biology. For an extra SGI UCAS point, individuals were between 1% (Computing) and 4% (Maths) more likely to take one of the STEM Highers. Individuals' SIMD quintile, their school's socio-economic composition and Higher cohort size were all insignificant.

The AH STEM subject choice models for the Twins/CSS' sub-sample data are shown in Table 3.11. The only significant variables are gender and individuals' Higher UCAS points.

Table 3.9

STEM Subject Choice at Standard Grade / Intermediate 2 : Twins/CSS All

	Maths	**Biology**	**Chemistry**	**Physics**	**Science**	**Computing**
Sex (M=1)	1.041	0.284***	0.952	3.650***	1.149	2.185***
SIMD 1	0.575***	0.445***	0.355***	0.423***	4.411***	0.750**
SIMD 2	0.739**	0.588***	0.465***	0.516***	3.721***	0.974
SIMD 3	0.756*	0.727***	0.670***	0.622***	2.357***	1.066
SIMD 4	0.837	0.825*	0.720***	0.818*	1.764**	0.881
SGI cohort size	1.004**	1.001	1.001	1.001	1.000	1.000
% pupils SIMD 1 & 2	0.977***	0.992***	0.990***	0.990***	0.999	0.994**
Observations	10872	10872	10872	10872	10832	10872
Log likelihood	-4504.0879	-6071.7272	-6236.9954	-5272.2307	-3047.9559	-5590.6432
Wald chi2(54)	496.45***	1003.97***	529.15***	1078.21***	380.50***	541.52***
Pseudo R^2	0.1570	0.1056	0.0566	0.1020	0.1244	0.0697

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (348 schools reducing to 345 for Science).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.10

STEM Subject Choice at Higher : Twins/CSS All

	Maths	**Biology**	**Chemistry**	**Physics**	**Computing**
Sex (M=1)	2.204***	0.331***	1.359***	7.306***	4.254***
SGI UCAS points	1.042***	1.014***	1.027***	1.026***	1.007***
SIMD 1	0.735	0.915	0.938	1.232	1.043
SIMD 2	1.046	0.836	0.944	0.971	0.786
SIMD 3	1.054	1.112	1.169	0.846	1.114
SIMD 4	0.866	0.900	1.109	1.197	0.916
Higher cohort size	1.001	1.002	1.001	0.999	0.998
% pupils SIMD 1 & 2	0.999	0.996	1.005	0.996	0.994
Observations	4347	4326	4347	4347	4324
Log likelihood	-1832.1545	-2067.3264	-1851.3102	-1612.0978	-1227.8767
Wald chi2(54, Biology/Computing 53)	1176.05***	753.17***	794.10***	865.54***	506.83 ***
Pseudo R^2	0.3862	0.1608	0.2199	0.2693	0.1496

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Individuals' SGI performance school quartile, Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (336 schools for Maths, Chemistry, Physics reducing to 334/333 for Biology/Computing).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.11

STEM Subject Choice at Advanced Higher : Twins/CSS All

	Maths	**Biology**	**Chemistry**	**Physics**	**Computing**
Sex (M=1)	2.708***	0.370***	1.123	6.894***	24.529***
Higher UCAS points	0.996*	1.002	0.994**	0.994**	0.994
SIMD 1	1.514	0.834	0.558	0.588	2.539
SIMD 2	0.672	0.723	0.610	0.835	3.891*
SIMD 3	0.791	1.402	0.824	0.722	1.242
SIMD 4	0.739	1.070	0.812	1.151	1.153
AH cohort size	1.007	0.998	0.991	0.994	1.026
% pupils SIMD 1 & 2	1.017*	0.995	1.001	1.001	1.028
Observations	1251	1250	1182	1213	917
Log likelihood	-542.58668	-477.62827	-453.13495	-298.20754	-125.7972
Wald chi2(50, Chem 49, Phy 48, Computing 41)	176.04***	141.73***	125.06***	159.31***	1775.01***
Pseudo R^2	0.1329	0.1119	0.1015	0.1813	0.3702

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Individuals' SGI UCAS points school quartile, Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (Clusters by subject: Maths 272, Biology 273, Chemistry 255, Physics 260, Computing 185).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Gender is significant for all the STEM subjects apart from Chemistry and exhibits the established pattern of uptake. Males are more likely to take AH Maths, Physics and Computing but less likely to take Biology compared to females. In keeping with the AH STEM choice models for the whole dataset (Table 3.8 above), individuals' total Higher UCAS points exerted a small negative influence on the uptake of Maths, Chemistry and Physics, reducing the likelihood of taking these subjects by 0.4–0.6% for one extra Higher UCAS point.

Whilst the results from the STEM subject choice models using the Twins/CSS subsample data would appear to mirror those for the population dataset, as noted above (Section 2.4), it is not possible to compare the results of logistic regressions directly across groups. Unobserved heterogeneity can vary between groups causing log-odds to be rescaled. To enable comparison, Tables 3.12, 3.13 and 3.14 show the average marginal effects of gender and socio-economic background on the probability of studying STEM subjects at SGI, Higher and Advanced Higher levels respectively for both the entire dataset and the Twins/CSS sub-sample.

At SGI level (Table 3.12), there is no gender effect on the probability of taking Maths as would be expected since its study is compulsory at this level. Where there is choice as to which STEM subjects are studied, the gendered nature of that choice is evident once again. The average marginal effects of gender on the probability of studying the different STEM subjects are broadly similar across the whole dataset and the Twins/CSS sub-sample, suggesting the latter to be representative of the population. On average, being male reduces the probability of taking Biology by 24-25% but increases the probability of taking Physics by approximately 20% and Computing by 13% for both the population and the Twins/CSS sub-sample. Being male has a small negative impact (1%) on the probability of taking Chemistry and a small positive impact on the probability of taking general Science (2%) for the population but this is not replicated for the Twins/CSS sub-sample. The established pattern of influence of socio-economic background on the uptake of STEM subjects found in the logistic regression models above is evident in the average marginal effects which are, again, similar in magnitude for the population data and the Twins/CSS sub-sample. The probability of taking a named Science, Maths or Computing at SGI level is reduced as individuals' SIMD falls sequentially compared to those from SIMD

5, the least deprived 20% of households. The effect varies from a 2% reduction in the probability of an individual from SIMD 4 taking Maths or Computing to a 20% reduction in the probability of an individual from SIMD 1 taking Chemistry (compared to individuals from SIMD 5). On the other hand, the probability of taking general Science is increased as SIMD falls, ranging from an average increase of 4-5% for those from SIMD 4 to an 11-12% increase for those from SIMD 1 (compared to SIMD 5 individuals).

The average marginal effects of gender on the probability of taking STEM subjects at Higher (Table 3.13), are similar for the population and the Twins/CSS sub-sample. Being male increases the probability of taking all STEM subjects apart from Biology; the effect varied between 4-6% for Chemistry and 23-24% for Physics, while the probability of taking Higher Biology was reduced by 15-17%. The influence of socio-economic background was seen to be much reduced at Higher in the subject choice logistic regression models above; SIMD was not significant at all for the Twins/CSS sub-sample and only displayed the usual pattern of impact for Maths and Physics for the population (Tables 3.10 and 3.7 respectively). The average marginal effects of SIMD on the probability of taking Maths and Physics, however, were very small (1-2%) and displayed no consistent pattern, suggesting that the influence of socio-economic background on STEM subject choice at Higher is negligible. An extra SGI UCAS point exerted a small, positive effect on individuals' probability of taking a STEM subject at Higher (between 0.1 to 0.6 of one percent, Computing and Maths respectively).

At Advanced Higher (Table 3.14), the average marginal effects of gender were reduced in magnitude for the named sciences across the population and for the Twins/CSS subsample, but not for Maths or Computing for the subsample. (As reported above, SIMD does not influence the uptake of STEM subjects at Advance Higher). An extra Higher UCAS point exerted a tiny, positive average marginal effect on the probability of taking an Advanced Higher STEM subject for the population (less than a tenth of one percent). For the Twins/CSS sub-sample, there were Higher UCAS points' effects of the same magnitude for Maths, Chemistry and Physics but not for Biology and Computing.

Table 3.12

STEM Subject Choice at Standard Grade / Intermediate 2 :

Average Marginal Effects of Gender and Socio-economic Background

Population						
SGI:	**Maths**	**Biology**	**Chemistry**	**Physics**	**Science**	**Computing**
Sex (M=1)	-0.003	-0.250***	-0.013***	0.211***	0.018***	0.139***
SIMD 1	-0.076***	-0.137***	-0.208***	-0.170***	0.105***	-0.064***
SIMD 2	-0.053***	-0.089***	-0.145***	-0.119***	0.086***	-0.032***
SIMD 3	-0.046***	-0.066***	-0.106***	-0.084***	0.062***	-0.019***
SIMD 4	-0.020*	-0.030***	-0.057***	-0.044***	0.035***	-0.018**
Twins/CSS Sub-sample						
SGI:	**Maths**	**Biology**	**Chemistry**	**Physics**	**Science**	**Computing**
Sex (M=1)	0.005	-0.238***	-0.010	0.204***	0.011	0.132***
SIMD 1	-0.073***	-0.153***	-0.201***	-0.136***	0.119***	-0.049**
SIMD 2	-0.040**	-0.100***	-0.149***	-0.104***	0.106***	-0.004
SIMD 3	-0.037*	-0.060***	-0.078***	-0.075***	0.069***	0.011
SIMD 4	-0.023	-0.036*	-0.064***	-0.032*	0.046**	-0.0214

Notes:
Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.13

STEM Subject Choice at Higher :

Average Marginal Effects of Gender and Socio-economic Background

Population					
Higher:	**Maths**	**Biology**	**Chemistry**	**Physics**	**Computing**
Sex (M=1)	0.114***	-0.154***	0 .066***	0.240***	0.127***
SGI points	0.006***	0.002***	0.006***	0.003***	0.001***
SIMD 1	-0.022***	-0.005	-0.013**	-0.011*	0.006
SIMD 2	-0.018***	-0.006	0.0015	-0.012**	0.008**
SIMD 3	-0.014*	0.002	0.001	-0.008*	0.011**
SIMD 4	-0.006	0.001	-0.004	-0.006	-0.000
Twins/CSS Sub-sample					
Higher:	**Maths**	**Biology**	**Chemistry**	**Physics**	**Computing**
Sex (M=1)	0.107***	-0.173***	0.042***	0.233***	0.120***
SGI points	0.006***	0.002***	0.004***	0.003***	0.001***
SIMD 1	-0.042	-0.014	-0.009	0.024	0.003
SIMD 2	0.006	-0.028	-0.008	-0.003	-0.020
SIMD 3	0.007	0.0166	.0215	-0.020	0.009
SIMD 4	-0.019	-0.017	.0142	0.021	-0.007

Notes:
Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Individuals' SGI UCAS points school quartile, Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.14

STEM Subject Choice at Advanced Higher :

Average Marginal Effects of Gender and Socio-economic Background

Population					
Adv Higher:	**Maths**	**Biology**	**Chemistry**	**Physics**	**Computing**
Sex (M=1)	0.116***	-0.098***	0.025***	0.146***	0.073***
Higher points	0.0004***	0.0002***	0.0005***	0.0004***	-0.0001***
SIMD 1	-0.012	-0.005	-0.051***	-0.024**	0.006
SIMD 2	-0.020**	-0.002	-0.022***	-0.010*	0.006*
SIMD 3	-0.005	0.006	-0.010*	-0.006	0.000
SIMD 4	0.001	-0.001	-0.011*	-0.002	-0.003
Twins/CSS Sub-sample					
Adv Higher:	**Maths**	**Biology**	**Chemistry**	**Physics**	**Computing**
Sex (M=1)	0.138***	-0.118***	0.014	0.136***	0.124***
Higher points	0.001*	0.000	0.001**	0.0004*	-0.000
SIMD 1	0.057	-0.021	-0.068	-0.038	0.036
SIMD 2	-0.055	-0.038	-0.058	-0.013	0.053
SIMD 3	-0.033	0.040	-0.023	-0.023	0.008
SIMD 4	-0.042	0.008	-0.025	0.010	0.006

Notes:
Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Individuals' SGI UCAS points school quartile, Local Authority, Urban/Rural location, Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5.3 STEM Subject Choice – The Influence of Testosterone

To investigate the influence of exposure to heightened levels of testosterone on STEM subject choices, the models were re-estimated for twin and CSS females only. Dummy variables were included to capture the socialisation effect of growing up with a male (as opposed to female) sibling and the effect of exposure to heightened testosterone levels over and above this. These were, respectively:

- css_msf = female from a closely spaced mixed-sex sibling pair
- twin_msf = female from a mixed-sex twin pair

The reference group comprised all remaining females from same sex twin and CSS pairs; that is females growing up alongside another female and with no exposure to heightened testosterone levels. The results at SGI and Higher levels are shown in Tables 3.15 and 3.16. The female models for Advanced Higher failed to converge.

The results for females alone at SGI level (Table 3.15) generally reflect those for the whole sub-sample. The effects of being a female with either a CSS brother or a twin brother were insignificant for all subjects. The same pattern of named and general science uptake associated with socio-economic status pertained, with the likelihood of studying a named science reducing (and that of general science rising) with increased levels of deprivation. The same broad effect for school socio-economic composition found in the models for the completer Twins/CSS sub-sample was repeated in the female models (the likelihood of taking a named science at SGI level falling by 1% for every 1% increase in the percentage of pupils from the most deprived 40% of households).

The average marginal effects of increased testosterone exposure and socio-economic background on female STEM subject choices at SGI can be seen in Table 3.17. These confirm that the more deprived the household an individual is from, the greater the reduction in the probability that they will be studying a named science at SGI level, and the greater the increase in the probability that they will take general Science compared to someone from an SIMD 5 household. The effects of SIMD on the probability of taking Maths, Chemistry and general Science in the Twins/CSS female models largely replicated those for the complete Twins/CSS sub-sample (Table 3.12). For an individual from SIMD 1 (compared to SIMD 5), these ranged from, approximately, a 4% reduction in the probability of taking Maths to a 20% reduction in the probability of taking Chemistry at SGI level and a 13% increase in the probability of taking general Science. Interestingly, the probability of taking SGI Biology is reduced by a further 3-4% in the Twins/CSS female only model for those from SIMDs 1 and 2 when compared with the general Twins/CSS Model (Table 3.12); minus -19% and minus -13% respectively compared to minus -15% and minus -10%. For SGI Physics, the opposite is seen with an approximate 3-4% reduction in the negative impact on the probability of uptake for individuals from SIMDs 1-3 in the Twins/CSS female only model (compared with the general Twins/CSS Model); for example, from minus -14% to minus -10% for SIMD 1 females.

At Higher, the range and magnitude of the effect of individuals' SGI points for the whole Twins/CSS sub-sample were mirrored in the female only models (Table 3.16), as confirmed by comparing the average marginal effects for both (Tables 3.13 and 3.17). An extra SGI UCAS point again exerted a small, positive effect on individuals' probability of taking a STEM subject at Higher (between 0.2 to 0.6 of one percent for Physics and Maths respectively). Those with a CSS brother were found to be more likely to take Higher Maths (43%). As noted above in Section 3.4.4, however, the female Biology, Physics and Computing models failed the goodness of fit test.

The female Twins/CSS AH STEM subject choice models failed to achieve convergence, probably due to the smaller sample size being drawn from a much lower national uptake at this level (see Figures 3.6 and 3.9 above). Choices made at AH level are subject to *double selection*. Individuals can only take a given subject at AH level if they have passed the Higher in that subject. Moreover, school provision of subjects at AH level (whether State or independent) is very narrow even within Scotland's largest education authority area, Glasgow City Council. It should be noted, however, that the really important STEM subject choices are made at Higher level and, before that, at SGI level. It is the subject choices made at Higher that determine subsequent entry to HE and degree type. In turn, access to STEM subjects at Higher in S5 (fifth year) are determined by subject choices made at SGI level and attainment therein.

Table 3.15

STEM Subject Choice at Standard Grade / Intermediate 2 : Twins/CSS Females

	Maths	**Biology**	**Chemistry**	**Physics**	**Science**	**Computing**
css_msf	0.950	0.957	0.961	1.016	0.881	0.871
twin_msf	0.959	1.013	0.996	1.022	1.094	0.810
SIMD 1	0.696^{*}	0.433^{***}	0.376^{***}	0.416^{***}	5.135^{***}	0.823
SIMD 2	0.752^{*}	0.573^{***}	0.467^{***}	0.532^{***}	4.444^{***}	0.932
SIMD 3	0.706^{*}	0.716^{**}	0.644^{***}	0.655^{***}	2.681^{***}	0.955
SIMD 4	0.803	0.861	0.756^{**}	0.747^{*}	2.301^{**}	0.857
SGI cohort size	1.004^{***}	1.000	1.002^{**}	1.001	1.001	1.000
% pupils SIMD 1 & 2	0.977^{***}	0.992^{**}	0.991^{***}	0.988^{***}	1.001	0.999
Observations	5531	5531	5531	5531	5484	5531
Log likelihood	2234.7991	-3526.5579	-3186.1721	-2019.4304	-1495.7013	-2440.3821
Wald chi2(51, Sci 49)	363.15***	401.63***	316.91***	215.30***	297.19***	257.31***
Pseudo R^2	0.1791	0.0698	0.0604	0.0550	0.1279	0.0526

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (343 schools, reducing to 333 for Science).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.16

STEM Subject Choice at Higher : Twins/CSS Females

	Maths	Biology	Chemistry	Physics	Computing
css_msf	1.428*	1.299	0.945	0.972	0.626
twin_msf	0.849	1.109	0.844	0.864	0.848
SGI UCAS points	1.044***	1.015***	1.029***	1.026***	1.008*
SIMD 1	0.714	1.195	0.970	1.619	0.479*
SIMD 2	0.990	0.913	1.027	1.016	0.461*
SIMD 3	1.219	1.046	1.206	1.389	1.338
SIMD 4	0.901	0.800	1.052	1.504	0.889
Higher cohort size	1.000	1.002	1.000	1.001	0.995
% pupils SIMD 1 & 2	1.000	0.992	1.001	0.993	0.993
Observations	2382	2371	2360	2348	1878
Log likelihood	-968.45089	-1262.6606	-974.31045	-642.39844	-381.67011
Wald chi2(55, 54, 53, 52, 43)	793.41***	512.28***	396.12***	325.38***	190.43***
Pseudo R^2	0.4016	0.1724	0.2368	0.2016	0.1546

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Individuals' SGI performance school quartile, Local Authority, Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (Clusters by subject: Maths 327, Biology 324, Chemistry 322 , Physics 320, Computing 252).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.17

STEM Subject Choice of Twin/CSS Females:

Average Marginal Effects of Increased Testosterone Exposure and Socio-economic Background

	Standard Grade / Intermediate 2					
	Maths	**Biology**	**Chemistry**	**Physics**	**Science**	**Computing**
css_msf	-0.007	-0.010	-0.008	0.002	-0.010	-0.019
twin_msf	-0.005	0.003	-0.001	0.002	0.007	-0.029
SIMD 1	-0.047*	-0.187***	-0.192***	-0.095***	0.127***	-0.027
SIMD 2	-0.037*	-0.125***	-0.149***	-0.068***	0.116***	-0.010
SIMD 3	-0.045*	-0.075**	-0.086***	-0.046***	0.076***	-0.006
SIMD 4	-0.028	-0.034	-0.055**	-0.031*	0.065**	-0.021

	Higher				
	Maths	**Biology**	**Chemistry**	**Physics**	**Computing**
css_msf	0.047*	0.047	-0.008	-0.002	-0.026
twin_msf	-0.021	0.019	-0.023	-0.012	-0.009
SGI points	0.006***	0.003***	0.004***	0.002***	0.0004*
SIMD 1	-0.044	0.032	-0.004	0.039	-0.040*
SIMD 2	-0.001	-0.016	0.004	0.001	-0.043*
SIMD 3	0.026	0.008	0.025	0.027	0.016
SIMD 4	-0.014	-0.040	0.007	0.033	-0.006

Notes:
Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Local Authority, Urban/Rural location and Year fixed effects and, for Highers' models, Individuals' SGI performance school quartile.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5.4 STEM Attainment – All Qualification Levels: Twins/CSS Data

Results for the STEM subject SGI and Higher attainment models for the whole Twins/CSS sub-sample are shown in Tables 3.18 and 3.20 respectively. There were convergence issues for some models. This was particularly the case at Higher with reduced sample sizes, whereby only a small number of observations were completely determined implying the existence of one-to-one mappings between the multinomial logit outcomes and certain independent variables. Investigation revealed the Local Authority dummies (32 of them) to be the cause of this at Standard Grade and, in addition, individuals' school quartile (based on their SGI points) at Higher and Advanced Higher. Accordingly, the LA fixed effects and school quartile variables were removed from all the Twins/CSS attainment models.

At SGI level (Table 3.18), sex was significant only for Maths (at the 5% level). Compared to failing, males were 22% more likely than females to achieve a low or middle pass in Maths. There were very few socio-economic background effects at the low pass level, evident only in Maths for SIMD quintile 1, Physics for SIMD quintile 3 and Computing for both SIMD quintiles 1 and 3; where the relative odds of obtaining a low pass in these subjects were reduced for these individuals compared to those from SIMD quintile 5. Socio-economic background effects increased in prevalence at the middle and high pass levels and generally exhibited the usual pattern of lower SIMDs being associated with increasingly lower relative odds of obtaining a given grade level. SIMD effects were least prevalent for general Science, where only individuals from the 20% most disadvantaged households (SIMD 1) were less likely to obtain a middle pass grade compared to those from the most affluent 20% of households (SIMD 5) while individuals from both SIMDs 1 and 2 were less likely to obtain a high pass. The SGI cohort size of schools was generally not significant. Schools' socio-economic composition was consistently significant at all levels of achievement for general Science only. Otherwise, it was significant for Maths at middle and high pass levels and Biology and Physics at the high pass.

Average marginal effects for the main variables of interest at SGI level are shown in Table 3.19. Being male reduced the probability of obtaining a high pass in a named science by some 4%-7% but increased the probability of achieving a low pass in Physics and a medium

Table 3.18

STEM Standard Grade/Intermediate 2 Attainment 2002-2009 : Twins/CSS Sub-Sample

(omitted category: Outcome 1 Fail)

	Maths	Biology	Chemistry	Physics	Science	Computing
Outcome 2: Pass low						
SEX (Male=1)	1.217*	0.882	0.843	1.591	0.933	0.715
SIMD 1	0.608*	0.595	0.884	0.643	0.648	0.252*
SIMD 2	0.784	0.728	0.906	0.585	0.875	0.356
SIMD 3	0.846	0.933	1.056	0.417*	0.853	0.313*
SIMD 4	0.991	1.272	1.503	0.901	1.695	0.468
SGI cohort size	1.001	1.000	0.998	1.000	0.995*	1.000
% pupils SIMD 1 & 2	0.996	1.002	1.001	0.996	0.982***	0.994
Outcome 3: Pass middle						
SEX (Male=1)	1.223*	0.908	1.047	1.180	0.818	0.735
SIMD 1	0.173***	0.207***	0.270***	0.273***	0.413*	0.069***
SIMD 2	0.321***	0.271***	0.329**	0.307***	0.478	0.159**
SIMD 3	0.424***	0.446*	0.646	0.361**	0.514	0.185**
SIMD 4	0.669	0.832	1.503	0.904	1.247	0.304
SGI cohort size	1.001	1.001	0.999	1.000	0.997	1.001
% pupils SIMD 1 & 2	0.993**	1.000	1.004	0.992	0.969***	0.993
Outcome 4: Pass high						
SEX (Male=1)	1.144	0.730	0.849	0.873	0.890	0.629*
SIMD 1	0.062***	0.067***	0.101***	0.139***	0.152*	0.033***
SIMD 2	0.129***	0.086***	0.151***	0.099***	0.196*	0.052***
SIMD 3	0.266***	0.290***	0.411*	0.205***	0.502	0.126***
SIMD 4	0.483***	0.598	1.217	0.732	0.922	0.250*
SGI cohort size	1.001	1.003	1.000	1.000	0.998	1.003
% pupils SIMD 1 & 2	0.988***	0.988*	0.993	0.984**	0.964**	0.994
Observations	8767	3444	3226	2507	1065	2625
Log pseudolikelihood	-10467.379	-3685.2907	-3250.5055	-2642.5852	-1085.1439	-2960.2595
Wald chi2(57)	897.98***	476.38 ***	371.46***	371.83***	242.29***	2491.41***
Pseudo R^2	0.0555	0.0637	0.0596	0.0600	0.0835	0.0563

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (Clusters by subject: Maths 345, Biology 337, Chemistry 332, Physics 335, Science 239, Computing 306).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.19

STEM Attainment at Standard Grade / Intermediate 2 :

Average Marginal Effects of Gender & Socio-economic Background for Twins/CSS Sub-sample

Variable	Grade	Maths	Biology	Chemistry	Physics	Science	Computing
Sex (M=1)	Fail	-0.014*	0.007	0.001	-0.006	0.014	0.014
	Low pass	0.010	0.002	-0.015	0.045**	0.014	-0.001
	Medium pass	0.011	0.028	0.049**	0.029	-0.028	0.014
	High pass	-0.007	-0.037*	-0.035*	-0.069**	0.000	-0.027
SIMD 1	Fail	0.101***	0.072***	0.043***	0.054**	0.074	0.096***
	Low pass	0.271***	0.166***	0.142***	0.109***	0.050	0.245***
	Medium pass	-0.132***	0.002	0.060	0.015	-0.084	-0.125**
	High pass	-0.240***	-0.241***	-0.245***	-0.179***	-0.040	-0.216***
SIMD 2	Fail	0.066***	0.061***	0.036**	0.056***	0.044	0.072**
	Low pass	0.207***	0.161***	0.120***	0.103***	0.098	0.187***
	Medium pass	-0.078***	0.018	0.041	0.102**	-0.103*	-0.010
	High pass	-0.195***	-0.240***	-0.197***	-0.260***	-0.038	-0.249***
SIMD 3	Fail	0.046**	0.034*	0.015	0.047**	0.040	0.064**
	Low pass	0.147***	0.108***	0.061**	0.028	0.068	0.095**
	Medium pass	-0.077***	-0.036	0.033	0.054	-0.098	-0.052
	High pass	-0.117***	-0.105***	-0.109***	-0.129***	-0.010	-0.107***
SIMD 4	Fail	0.020	0.007	-0.011	0.006	-0.050	0.044
	Low pass	0.090***	0.068*	0.009	0.006	0.095	0.075*
	Medium pass	-0.037*	-0.003	0.041	0.030	-0.032	-0.052
	High pass	-0.074***	-0.072**	-0.039	-0.043	-0.013	-0.067*

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

pass in Chemistry by 5%. In general, for individuals from lower SIMDs there was an increased probability of obtaining a low pass or failing and a reduced probability of achieving a medium or high pass compared to those from SIMD 5. The lower an individual's SIMD, the greater was the impact. Compared to those from SIMD 5, on average, individuals from SIMD 1 had an increased probability of failing a STEM SGI of between 4% and 10% (Chemistry and Maths), and a reduced probability of gaining a high pass of between 18% and 25% (Physics and Chemistry). These results were broadly in line with those for the population (Table 2.6, p119), suggesting that the Twins/CSS sub-sample was representative.

There was almost a complete absence of significant average marginal effects for general Science. The relatively low influence of SIMD effects on attainment in general Science, compared to subject choice may reflect the fact that this has tended to be taken by academically weaker pupils and/or those who are not interested in science subjects (but have to take such because of curriculum structure).

At Higher, for the whole Twins/CSS' sub-sample models (Table 3.20), the only variable that was consistently significant for all the STEM subjects, at all pass levels, was individuals' prior attainment/ability as measured by their SGI UCAS points. Its impact increased from making a low grade pass, in each subject, some 2%-3% more likely for an extra SGI UCAS point (compared to failing) to making a middle grade pass 4%-6% more likely and a top grade pass 6%-10% more likely. Being male was significant for Computing at low pass, Maths and Chemistry at middle and high pass, and Physics at high pass. Boys were only 36% as likely as girls to obtain a low pass in Computing (compared to failing) but more likely to obtain middle passes in Maths (83%) and Chemistry (71%). They were 77% more likely to achieve a high pass in Maths and twice as likely to obtain this top grade in Chemistry and Physics. Individuals' socio-economic background was rarely significant and there was no evidence of any consistent pattern in terms of deprivation levels. Schools' Higher cohort size and socio-economic composition were not significant in general with the notable exception of middle and high pass in Physics, where a 1% increase in the proportion of pupils from the most deprived 40% of households (% pupils SIMD 1 & 2) reduced the likelihood of these by 1.5%-3% respectively (compared to failing).

Table 3.21 gives the average marginal effects for the main variables of interest at Higher for the Twins/CSS sub-sample. For Maths, being male, as opposed to female, reduced the probability of failing by 6% and increased the probability of securing a middle or high pass (5% and 4% respectively). Additionally, being male increased the probability of achieving a high pass in Chemistry and Physics by 7%. On average, the probability of failing Higher Computing increased by 8% for males compared to females and the probability of securing a low pass in either Computing or Chemistry was reduced (16% and 5% respectively). One extra SGI UCAS point was seen to reduce the probability of failing or low pass (0.2 and 0.5 of one percent) and increase the probability of achieving a middle or high pass (0.1 and 0.7 of one percent). Again, these effects were in line with those for the population as a whole (Table 2.14). For individuals from SIMD 1, the most deprived 20% of households, the probability of failing Higher Maths was increased by 12% compared to those from SIMD 5 (the least deprived 20% of households); an effect three times as large as that for the population (Table 2.14). In general, socio-economic background effects were sparse with no consistent pattern.

Table 3.20
STEM Higher Attainment 2002-2009 : Twins/CSS Sub-sample
(omitted category: Outcome 1 Fail)

	Maths	Biology	Chemistry	Physics	Computing
Outcome 2: Pass low					
SEX (Male=1)	1.193	1.214	1.085	1.236	0.362**
SGI UCAS points	1.017***	1.033***	1.017***	1.016***	1.014**
SIMD 1	0.453**	0.754	0.523	1.660	0.855
SIMD 2	0.754	1.534	0.729	1.252	2.162
SIMD 3	0.803	1.066	0.810	1.363	1.079
SIMD 4	0.919	0.982	0.578*	1.429	1.176
Higher cohort size	0.995*	0.997	1.007*	0.998	1.009
% pupils SIMD 1 & 2	0.993	0.990	1.002	0.988	1.001
Outcome 3: Pass middle					
SEX (Male=1)	1.828***	0.911	1.709*	1.513	0.815
SGI UCAS points	1.038***	1.063***	1.045***	1.037***	1.044***
SIMD 1	0.408**	1.335	0.862	0.894	1.666
SIMD 2	0.851	1.377	0.924	0.827	1.184
SIMD 3	1.046	1.869	1.137	1.231	0.816
SIMD 4	1.310	1.774	0.690	1.338	1.130
Higher cohort size	1.000	0.995	1.000	1.000	1.015*
% pupils SIMD 1 & 2	1.000	0.989	0.991	0.985*	1.007
Outcome 4: Pass high					
SEX (Male=1)	1.774***	1.584	2.217**	2.077**	1.190
SGI UCAS points	1.060***	1.097***	1.077***	1.072***	1.079***
SIMD 1	0.463	0.312	0.487	1.819	0.697
SIMD 2	0.655	0.544	0.737	1.173	1.420
SIMD 3	0.603*	0.918	0.532	1.150	0.633
SIMD 4	0.946	0.989	0.754	2.544**	1.801
Higher cohort size	0.997	0.990*	0.999	0.996	1.006
% pupils SIMD 1 & 2	0.998	0.990	0.990	0.973***	1.006
Observations	1926	1110	1024	892	450
Log pseudolikelihood	-2270.389	-1132.8961	-1129.6066	-984.30238	-461.75047
Wald chi2(60)	303.72***	442.15***	249.62***	279.52***	225.93***
Pseudo R^2	0.1485	0.2633	0.2021	0.1962	0.2586

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant. SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households. Other independent variables included: Urban/Rural location and Year fixed effects. Cluster-robust standard errors at the school level (Clusters by subject: Maths 304, Biology 281, Chemistry 268, Physics 266, Computing 177). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.21

STEM Attainment at Higher :

Average Marginal Effects of Gender and Socio-economic Background for Twins/CSS Sub-sample

Variable	Grade	Maths	Biology	Chemistry	Physics	Computing
Sex (M=1)	Fail	-0.059**	-0.017	-0.043	-0.050	0.080*
	Low pass	-0.031	0.020	-0.053*	-0.021	-0.156***
	Medium pass	0.053**	-0.058	0.027	0.003	0.013
	High pass	0.037*	0.055	0.069**	0.069*	0.063
SGI UCAS points	Fail	-0.005***	-0.006***	-0.004***	-0.004***	-0.004***
	Low pass	-0.002***	-0.002***	-0.003***	-0.002***	-0.002***
	Medium pass	0.001***	0.001***	0.001*	0.000	0.001**
	High pass	0.006***	0.006***	0.006***	0.007***	0.005***
SIMD 1	Fail	0.120**	0.026	0.063	-0.042	-0.004
	Low pass	-0.049	-0.017	-0.068	0.057	-0.044
	Medium pass	-0.056	0.143*	0.068	-0.092	0.117
	High pass	-0.015	-0.152*	-0.063	0.077	-0.069
SIMD 2	Fail	0.040	-0.034	0.030	-0.010	-0.071
	Low pass	-0.018	0.080*	-0.035	0.037	0.103**
	Medium pass	0.019	0.068	0.029	-0.057	-0.039
	High pass	-0.041	-0.115*	-0.024	0.029	0.006
SIMD 3	Fail	0.029	-0.024	0.022	-0.031	0.012
	Low pass	-0.014	-0.021	-0.015	0.031	0.037
	Medium pass	0.057*	0.102**	0.085*	0.009	-0.008
	High pass	-0.072*	-0.056	-0.091**	-0.009	-0.041
SIMD 4	Fail	-0.005	-0.017	0.056	-0.059	-0.028
	Low pass	-0.026	-0.035	-0.058	-0.002	0.002
	Medium pass	0.055*	0.093**	-0.011	-0.042	-0.027
	High pass	-0.025	-0.040	0.013	0.103**	0.054

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5.5 STEM Attainment – The Influence of Testosterone

Results from the female-only Twins/CSS models at SGI and Higher are shown in Tables 3.22 and 3.24 respectively. As above (Section 3.5.4), these models were estimated without LA fixed effects and, at Higher, school quartile variables to mitigate convergence issues. In keeping with the female STEM subject choice models, the attainment models included dummy variables for being a female from a mixed-sex CSS pair (css_msf) and being a female twin from a mixed-sex pair (twin_msf). The former captures any socialisation effect from being brought up with a male while the latter captures any testosterone effect over and about this. As before, the reference group comprised other females from same-sex CSS or twin pairs; ie those not exposed to heightened testosterone levels and/or subject to any *male* socialisation effects.

In the event, there was no evidence for any testosterone effect; the dummy variable for being a female twin from a mixed-sex pair (twin_msf) was insignificant for all the STEM subjects at both SGI and Higher. No consistent socialisation effects were found.

At SGI level (Table 3.22), CSS females with brothers (css_msf) were found to be less likely than the control group to achieve a middle pass in general Science (compared to failing) and less likely to obtain a high pass in either general Science or Biology. Socio-economic background effects were completely absent in explaining the likelihood of a low pass. They became consistently important in Maths, Biology and Computing at middle and high pass grades and Physics at high pass, essentially exhibiting the usual pattern. Females from SIMD quintiles 1 and 2 were only 6% as likely as those from SIMD 5 to obtain a high pass in general Science rather than fail. SGI cohort size effects were absent while schools' socio-economic composition reduced the likelihood of obtaining a low or middle pass (compared to failing) in general Science by 2-3% s and a high pass in Maths by 1%.

The accompanying average marginal effects for the main variables of interest at SGI level for the Twins/CSS female only models are shown in Table 3.23. There is evidence of a negative socialisation effect for CSS females with a brother (css_msf) whereby, on average, the probability of their achieving a high pass in Maths, Biology, Physics or general Science is reduced by between 4% (Maths) to 43% (general Science) compared to the control group. They had an increased probability of obtaining a low pass in Maths, Physics

Table 3.22

STEM Standard Grade/Intermediate 2 Attainment 2002-2009 : Twins/CSS FEMALES

(omitted category: Outcome 1 Fail)

	Maths	Biology	Chemistry +	Physics	Science	Computing
Outcome 2: Pass low						
css_msf	1.122	0.620	1.403	2.696	0.655	1.109
twin_msf	1.161	0.830	1.177	3.887	1.337	1.693
SIMD 1	0.642	0.647	0.547	0.461	0.549	0.227
SIMD 2	0.781	0.821	0.751	0.531	0.807	0.799
SIMD 3	1.008	1.026	0.657	0.447	0.850	0.168
SIMD 4	0.889	1.975	1.095	0.791	1.212	0.459
SGI cohort size	1.001	1.001	0.998	0.995	0.997	1.000
% pupils SIMD 1 & 2	0.998	1.006	1.000	0.999	0.980**	0.984
Outcome 3: Pass middle						
css_msf	0.956	0.795	1.506	1.082	0.342*	0.733
twin_msf	1.151	0.935	1.334	3.485	0.663	1.119
SIMD 1	0.149***	0.207***	0.246	0.311	0.289	0.056**
SIMD 2	0.297***	0.315**	0.331	0.269	0.241	0.324
SIMD 3	0.467**	0.505	0.398	0.359	0.241	0.125*
SIMD 4	0.643	1.228	1.308	1.177	0.534	0.276
SGI cohort size	1.001	1.002	0.999	0.996	1.000	1.002
% pupils SIMD 1 & 2	0.997	1.004	1.000	0.992	0.967***	0.988
Outcome 4: Pass high						
css_msf	0.779	0.539*	0.678	0.683	0.000***	0.606
twin_msf	1.153	0.836	1.290	3.403	2.388	1.467
SIMD 1	0.064***	0.082***	0.097	0.176*	0.064*	0.022***
SIMD 2	0.145***	0.105***	0.172	0.142**	0.061*	0.097*
SIMD 3	0.296***	0.311**	0.267	0.232*	0.219	0.082*
SIMD 4	0.459**	0.819	0.971	1.490	0.484	0.208
SGI cohort size	1.002	1.004	1.000	0.995	1.002	1.004
% pupils SIMD 1 & 2	0.987**	0.989	0.988	0.986	0.980	0.994
Observations	4457	2423	1674	719	518	975
Log pseudolikelihood	-5324.8075	-2576.739	-1679.2442	-699.82155	-510.48652	-1069.0752
Wald chi2(60, Chem 47, Sci 147)	588.76***	351.71***	-	3775.26***	5500.72***	2923.29***
Pseudo R^2	0.0643	0.0697	0.0765	0.1096	0.1127	0.0801

Notes: Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles. Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Urban/Rural location and Year fixed effects. Cluster-robust standard errors at the school level (Clusters by subject: Maths 339, Biology 327, Chemistry 313, Physics 243, Science 186, Computing 261). $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ $^{+}$ Variance matrix is nonsymmetric or highly singular

Table 3.23

STEM Attainment at Standard Grade / Intermediate 2 :

Average Marginal Effects of Increased Testosterone Exposure and Socio-economic Background for Twins/CSS Females

Variable	Grade	Maths	Biology	Chemistry [94]	Physics	Science	Computing
CSS_msf	Fail	0.000	0.017	-0.005	-0.003	0.123**	0.007
	Low pass	0.044*	-0.023	0.022	0.089***	0.280***	0.084
	Medium pass	-0.006	0.066*	0.139	0.035	0.022	-0.032
	High pass	-0.038*	-0.060*	-0.156	-0.121*	-0.425***	-0.059
Twin_msf	Fail	-0.012	0.005	-0.008	-0.050	-0.010	-0.011
	Low pass	0.007	-0.013	-0.010	0.018	0.113	0.063
	Medium pass	0.003	0.024	0.018	0.025	-0.128*	-0.075
	High pass	0.002	-0.016	0.000	0.007	0.025	0.023
SIMD 1	Fail	0.111***	0.067***	0.046	0.051	0.102	0.090*
	Low pass	0.286***	0.171***	0.108	0.041	0.075	0.275***
	Medium pass	-0.170***	-0.024	0.074	0.051	-0.121	-0.105
	High pass	-0.227***	-0.214***	-0.228	-0.143	-0.057	-0.260***
SIMD 2	Fail	0.072***	0.052***	0.035	0.056	0.078	0.037
	Low pass	0.200***	0.157***	0.102	0.066	0.192	0.222***
	Medium pass	-0.100***	0.023	0.034	0.046	-0.208**	0.008
	High pass	-0.171***	-0.232***	-0.171	-0.168**	-0.062*	-0.267***
SIMD 3	Fail	0.039	0.027	0.029	0.045	0.070	0.073
	Low pass	0.161***	0.107***	0.061	0.024	0.185	0.052
	Medium pass	-0.082**	-0.020	0.016	0.040	-0.230*	-0.019
	High pass	-0.117***	-0.114***	-0.105	-0.109*	-0.026	-0.105*
SIMD 4	Fail	0.028	-0.011	-0.005	-0.007	0.010	0.041
	Low pass	0.071**	0.081**	-0.007	-0.039	0.145	0.094
	Medium pass	-0.027	0.014	0.064	-0.017	-0.138	-0.047
	High pass	-0.072***	-0.084***	-0.052	0.062	-0.017	-0.087

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable. SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[94] No standard errors computed as variance matrix is nonsymmetric or highly singular.

or Science, ranging from 4% (Maths) to 28% (general Science) but also an increased probability of failing general Science of 12%. Socio-economic background effects were similar in pattern and magnitude to those for the whole Twins/CSS sub-sample in Maths and Biology (as shown above in Table 3.19). Individuals from lower SIMDs had an increased probability of obtaining a low pass or failing and a reduced probability of achieving a middle or high pass compared to those from SIMD 5. Socio-economic background effects were largely absent for Physics and Science, and sporadic for Computing (and could not be calculated for Chemistry).

Table 3.24 provides results for the female-only Twins/CSS models for Maths and the named Sciences' Highers as the model for Higher Computing could not be estimated reliably. [95] A socialization effect is evident for CSS females with brothers (css_msf) at the high pass level in all subjects. CSS females with brothers were between only 18% (Physics) and 33% (Maths) as likely as the control group females (those with female siblings) to achieve this grade compared to failing. In common with the population data and Twins/CSS whole sub-sample models (Tables 2.12 and 3.20), SGI UCAS points were significant for all subjects, at all levels, with their impact of increasing the likelihood of obtaining a given level of pass varying between 2% (Maths/Chemistry low pass) and 11% (Physics high pass) for an additional point. Socio-economic background effects were absent apart from Physics at high pass, where small sample numbers with outliers produced an odds ratio suggesting that those from SIMD 4 were over four times as likely to achieve this as those from SIMD 5 (compared to failing). Higher cohort size and schools' socio-economic composition effects were largely absent also apart from Biology and Physics at high pass. An increase in the size of a school's Higher cohort by one percent, reduced the relative odds of obtaining a high pass in Biology and Physics by one percent and two percent respectively. A one percent increase in proportion of a school's pupils from SIMDs 1 and 2 was seen to reduce the likelihood of a high pass in Physics by 5%.

Average marginal effects for the main variables of interest at Higher for the Twins/CSS female-only models are shown in Table 3.25. There were a number of significant effects for being a CSS female with a brother (css_msf) that, while exhibiting no particularly

[95] Only one observation was completely determined.

consistent pattern, indicated that these girls tended to perform better in Computing as opposed to Maths and Science compared to the control group. They had a reduced probability of failing Higher Computing (22%) but an increased probability of failing Maths and Chemistry (8% and 9%). Their probability of achieving a high pass in Computing was increased by 23% while their probability of securing such for in Maths and Biology was reduced by 13% and 16% respectively. There were just three significant effects for twin girls from mixed-sex pairs (twin_msf), these were in Chemistry and Computing. Their probability of obtaining a middle pass in Chemistry was increased by 10% compared to the control group, however their probability of achieving a high pass in the subject was reduced by 10% while their probability of a high pass in Computing was decreased by 25%. To the extent that the last two results provide any evidence of a testosterone transfer effect, they are counter-intuitive suggesting that females from mixed-sex twin pairs tend to perform less well in these two STEM subjects compared to the control group. The impact of an extra SGI UCAS point exhibited the same pattern as for the whole Twin/CSS sub-sample with similar magnitudes; in general reducing the probability of failing or a low pass (0.2 and 0.5 of one percent) and increasing the probability of gaining a middle or high pass (0.1 and 0.9 of one percent). Socio-economic background effects were again sparse and sporadic at Higher, exhibiting no consistent pattern but, where significant, these tended to be for the same qualifications, at the same grades and similar magnitudes, as for the whole Twin/CSS sub-sample. For instance, in the female-only Twins/CSS model, the probability of failing Higher Maths increased by 10% for those from SIMD 1 compared to 12% for SIMD 1 individuals in the whole Twin/CSS sub-sample.

AH STEM attainment models were estimated but convergence was not consistent across data cuts or subjects. For Twin/CSS females only, none of the subject attainment models achieved convergence.

Table 3.24

STEM Higher Attainment 2002-2009 : Twins/CSS FEMALES

(omitted category: Outcome 1 Fail)

	Maths	**Biology**	**Chemistry**	**Physics**
Outcome 2: Pass low				
css_msf	0.578	0.956	0.499	0.710
twin_msf	0.760	0.895	0.610	0.442
SGI UCAS points	1.019***	1.031***	1.020***	1.028**
SIMD 1	0.552	0.730	0.600	1.684
SIMD 2	0.709	1.349	0.686	1.601
SIMD 3	0.958	1.040	0.821	3.289
SIMD 4	0.784	1.118	0.749	4.080
Higher cohort size	0.998	0.998	1.008	0.991
% pupils SIMD 1 & 2	0.993	0.990	0.993	0.986
Outcome 3: Pass middle				
css_msf	0.764	0.591	0.551	0.348
twin_msf	0.715	0.876	1.088	0.800
SGI UCAS points	1.047***	1.062***	1.060***	1.060***
SIMD 1	0.519	1.158	1.511	0.512
SIMD 2	0.818	1.184	1.011	0.559
SIMD 3	1.085	1.719	1.178	4.903
SIMD 4	1.071	1.657	0.859	3.026
Higher cohort size	1.001	0.995	1.000	0.991
% pupils SIMD 1 & 2	0.996	0.987	1.000	0.993
Outcome 4: Pass high				
css_msf	0.329**	0.188**	0.212*	0.184*
twin_msf	0.666	0.637	0.447	1.104
SGI UCAS points	1.067***	1.100***	1.085***	1.107***
SIMD 1	0.315	0.298	0.524	1.675
SIMD 2	0.659	0.415	0.460	1.006
SIMD 3	0.795	0.735	0.449	1.991
SIMD 4	0.637	0.996	0.514	4.464*
Higher cohort size	1.000	0.989*	1.000	0.978**
% pupils SIMD 1 & 2	1.001	0.991	0.997	0.954**
Observations	994	815	546	254
Log pseudolikelihood	-836.66968	-784.68404	-565.30921	-243.90679
Wald chi2(63, 61 Comp)	274.26***	384.98***	211.00***	250.34***
Pseudo R^2	0.1707	0.2593	0.2512	0.2899

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Other independent variables included: Urban/Rural location and Year fixed effects.
Cluster-robust standard errors at the school level (Clusters by subject: Maths 262, Biology 258, Chemistry 210, Physics 130).
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.25

STEM Attainment at Higher :

Average Marginal Effects of Increased Testosterone Exposure & Socio-economic Background for Twins/CSS Females

Variable	Grade	Maths	Biology	Chemistry	Physics
CSS_msf	Fail	0.084*	0.041	0.090*	0.075
	Low pass	-0.023	0.087	-0.008	0.080
	Medium pass	0.066	0.033*	0.054	-0.013
	High pass	-0.127**	-0.161**	-0.136	-0.143
Twin_msf	Fail	0.045	0.018	0.041	0.040
	Low pass	-0.009	0.006	-0.047	-0.106
	Medium pass	-0.011	0.017	0.099*	-0.003
	High pass	-0.025	-0.041	-0.093*	0.069
SGI UCAS points	Fail	-0.005***	-0.005***	-0.005***	-0.005***
	Low pass	-0.002***	-0.002***	-0.004***	-0.004***
	Medium pass	0.002***	0.001***	0.002***	0.000
	High pass	0.006***	0.006***	0.006***	0.009***
SIMD 1	Fail	0.103*	0.032	0.029	-0.014
	Low pass	-0.010	-0.018	-0.076	0.083
	Medium pass	0.009	0.131	0.141*	-0.177
	High pass	-0.102	-0.144*	-0.094	0.107
SIMD 2	Fail	0.045	-0.017	0.036	-0.005
	Low pass	-0.027	0.074	-0.026	0.094
	Medium pass	0.015	0.073	0.079	-0.118
	High pass	-0.033	-0.129*	-0.089	0.028
SIMD 3	Fail	0.007	-0.018	0.020	-0.109
	Low pass	0.001	-0.016	-0.003	0.054
	Medium pass	0.033	0.111*	0.095*	0.139
	High pass	-0.040	-0.077	-0.112*	-0.084
SIMD 4	Fail	0.029	-0.026	0.034	-0.120*
	Low pass	-0.022	-0.012	-0.008	0.073
	Medium pass	0.057	0.079*	0.040	-0.016
	High pass	-0.064	-0.041	-0.066	0.063

Notes: Average marginal effects measure the average change in the probability that y=1, i.e. the probability that an individual takes a STEM subject at the level, for a one unit increase in a given independent variable.
SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles.
Omitted SIMD reference category = SIMD 5 – the least deprived (most affluent) 20% of households.
Parent regression models included: Urban/Rural location and Year fixed effects. Cluster-robust standard errors at the school level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.6 Conclusions

STEM knowledge and skills are of critical importance to the Scottish and UK economies in terms of economic growth and employment. Jobs requiring STEM skills are expected to increase at twice the rate of other occupations over the next decade for the UK as a whole. The rapid increase in STEM related jobs in Scotland since 2010, particularly in professional and/or IT occupations, is forecast to continue but, at the same time, skill shortages are set to worsen and, therefore, may act as a brake on STEM-driven economic growth and employment. The growth in STEM jobs is to be welcomed as the sector provides relatively well-paid employment opportunities, however, in terms of access and participation, there are clear equity and social inclusion issues. STEM sector employment is highly gendered with women being over-represented in the health sector (medicine, dentistry and general health care jobs) and men over-represented in architecture, engineering and construction. Gender segregation in STEM employment reflects the uptake of degree programmes at university, with women outnumbering men in medicine and dentistry and men outnumbering women in engineering and architecture, and, before that, the uptake of STEM subjects in secondary school that directly determines HE entry and course type. Established, persistent gender segregation exists in the choice of STEM subjects for formal examination and qualification at secondary school with Biology being the female science of choice whilst males are far more likely to opt for Physics and, latterly, Computing. Additionally, individuals from more deprived backgrounds are less likely to choose to study STEM subjects in secondary school with subsequent consequences for their future employment and earnings' opportunities. With both economic growth and equity issues in mind, the Scottish Government's STEM Education and Training Strategy simultaneously aims to increase the STEM capability and skills of individuals at all levels of education and redress the gender and socio-economic imbalance in STEM education and employment. Specific targets for 2022 include: a 10% increase Mathematics' passes at National 5 (previously, Credit Standard Grade or Intermediate 2), a reduction in the attainment gap in Higher STEM subjects between those from the least deprived and the most deprived SIMD quintiles to 25 points and an increase in the number of females passing Physics by 15% and Computing by 20%.

Much research, particularly in the field of Behavioural Genetics, suggests that educational achievement is moderately (approximately 40%) to highly heritable (70% or more) in reading and numeracy with science perhaps less so. It has been postulated that the more equitable is the education system, the more genetic differences appear to account for differences in attainment between individual children. In this respect, it has been argued that greater heterogeneity in the teaching of science, as opposed to the more standardized curriculums and lessons for reading and Maths, may explain the apparent lower heritability of attainment in science. General cognitive ability is regarded as the main determinant of educational attainment and there is no evidence of gender differences. Despite this, clear, persistent, average mean gender differences exist with respect to numeracy-based and language-based subject attainment, with males tending to perform better in the former and females the latter.

This chapter explored whether gendered STEM subject choice and attainment in Scottish secondary schools might have a genetic component in addition to being socially constructed as is generally accepted to be the case. Specifically, the potential impact of testosterone on STEM subject choice and attainment was explored by exploiting the Twin TTT hypothesis. The TTT hypothesis postulates that female co-twins from mixed-sex twin pairs maybe exposed to heightened levels of testosterone in-utero. There is evidence to suggest that females exposed to higher levels of testosterone exhibit more masculine physical, behavioural and cognitive traits. If there is a genetic component to STEM subject choice and attainment, it is reasonable to expect that this might be more in evidence in the context of the acknowledged equity of provision of the Scottish state education system and uncovering such should help to inform or improve STEM educational policy.

For the analysis, SQA administrative data were rendered to produce a sub-sample dataset of twins and closely spaced siblings (CSS), identified through their dates of birth and common households. Use of the CSS group was designed to provide a control for similarity of socialisation, in terms of the different effects of growing up with a same or opposite sex sibling, in an attempt to identify any separate testosterone effect on the STEM choices and attainment of female co-twins from mixed-sex pairs. Choice of and attainment in all STEM subjects was examined at all qualification levels; Standard Grade/Intermediate 2 (SGI), Higher (H) and Advanced Higher (AH). Binary logistic regression was used to model STEM subject choice and, for comparison purposes, this was carried out for the full population data as well as for the Twins/CSS sub-sample data. Attainment was modelled using multinomial logit analysis as, in general, the original ordered logit models failed the test of the proportional odds' assumption. For both sets of analyses, models were estimated using the entire Twins/CSS sub-sample data and then separately for females to investigate whether female co-twins, (potentially) exposed to higher levels of testosterone, differed from other females in their STEM subject choices and attainment.

For SGI STEM subject choices, results from the Twins/CSS All and Female models reflected those for the population as a whole. As expected, there were clear patterns of subject gender segregation with, approximately, females being three times more likely to choose Biology than males, males three times more likely to choose Physics and twice as likely to opt for Computing. Males were, however, some 25% more likely to choose general Science than females; confirmation perhaps of the greater variation in male educational performance that has been found elsewhere as general science tends to be taken by weaker students. On average, at SGI level, being male reduced the probability of choosing Biology by approximately 25% and increased the probability of taking Physics by 20% and Computing by 13% compared to being female. There was a small positive increase in the probability of taking general Science (1%-2%) for males. There was no indicative evidence of any TTT impact on the choices of female co-twins from mixed-sex twin pairs. There was, however, stark evidence of the influence of socio-economic background on STEM subject choice at this level with the likelihood of taking a named science reducing, and that of taking general Science rising, sequentially, as individuals' SIMD backgrounds went from least deprived (SIMD 5) to most deprived (SIMD 1). For instance, individuals from SIMD 1, the most deprived 20% of households, were only 35% as likely to take SGI Chemistry or Physics as those from the least deprived (most affluent) 20% of households but were almost four times as likely to take general Science (Table 3.6). On average, for individuals from SIMD 1, compared to those from SIMD 5, the probability of taking general Science was increased by 11% and the probability of taking other STEM subjects reduced by between 5% to 20%.

The gendered pattern of subject choice was repeated at Higher with males seen to be more likely to take all STEM subjects apart from Biology in both the population and Twins/CSS sub-sample models. On average, at Higher, being male reduced the probability of taking Biology by 15-17% and increased the probability of taking Physics by 23-24%. Socio-economic background effects were much reduced and only consistently present for Maths and Physics in the models using the population data, indicating that individuals were less likely to choose these subjects for study at Higher, the lower their household SIMD. The average marginal effects were small with a 1-2% decrease in the probability of taking Maths and/or Physics for individuals from SIMDs 1 to 3 compared to those from SIMD 5. On the other hand, the likelihood of choosing Computing at Higher was seen to be greater for those individuals from SIMDs 2 and 3 than for those from SIMD 5 with a small increase in the probability of taking such of 1%. The only variable that was consistently significant across all subjects at Higher and across all datasets was an individual's previous attainment/ability as measured by their total SGI UCAS points. The impact of this ranged from between a 1% and 4% increase in the odds of taking a given STEM Higher for a one point increase in an individual's SGI UCAS points. This translated into a small (less than 1%), positive increase in the probability of taking a STEM Higher for one extra SGI UCAS point. Once again, there was no evidence of any TTT impact on the choices of female co-twins from mixed-sex twin pairs. The AH level models for the female-only Twins/CSS sub-sample did not achieve convergence. Those for the population data and whole Twins/CSS sub-sample reflected the same gendered pattern of subject choice as at Higher, with males more likely to choose all STEM subjects apart from Biology. On average, being male reduced the probability of taking AH Biology by some 10-12% and increased the probability of taking other STEM subjects by between 3% (Chemistry) and 15% (Physics). SIMD effects were only evident for Chemistry and to a lesser extent Physics in the population models with, for instance, on average the probability of an individual from SIMD 1 taking AH Chemistry or Physics being reduced by 5% and 2% respectively.

STEM SGI attainment multinomial logit models, estimated for the whole Twins/CSS sub-sample and then separately for female Twins/CSS, revealed only the occasional gender or gender-related (twin or CSS) effect but no evidence for any TTT impact on the attainment of female co-twins from mixed-sex twin pairs. Individuals' socio-economic background was found not to be relevant at the low pass levels apart from Computing in the Twins/CSS All model (although there was no consistent pattern here). SIMD was found to grow in importance, however, as attainment levels rose, with those from more deprived backgrounds being significantly less likely to achieve middle and high STEM SGI passes. On average, for the whole Twins/CSS sub-sample, being from an SIMD 1 household (as opposed to SIMD 5) reduced the probability of a high pass by 18% -25% in STEM subjects at SGI level apart from general Science, where socio-economic background was not significant. Average marginal effects of similar magnitude were found at high pass for Maths, Biology and Computing in the female model.

For the Higher STEM attainment models (similar to the Higher choice models), only individuals' SGI UCAS points were significant for all subjects, at all pass levels, in both sets of Twins/CSS models (whole sub-sample and females only). The impact of an extra SGI UCAS point varied between a 2% increase in the likelihood of obtaining a low pass (as opposed to a fail) across all subjects and a 10% increase in the likelihood of gaining a high pass in Biology. On average, the impact of an extra SGI UCAS point increased the probability of obtaining a middle or high pass by between 0.001 and 0.009 of one percent. SIMD effects were almost non-existent in all models; where they did occur, there was no consistent pattern. Males were seen to be more likely to achieve middle and high passes in Maths and Chemistry, and high passes in Physics than females (whole Twins/CSS sub-sample models). CSS females with brothers were significantly less likely to achieve a high pass in any of the STEM subjects compared to other females. Aside from this, there were no other gender-related effects and so no evidence of any TTT impact on the attainment of female co-twins from mixed-sex twin pairs.

To conclude, whilst subject choices made at Higher level are crucial as these determine HE entry, course type and future occupation, these choices are fed by SGI, now National 5, level choices and attainment. Whilst STEM subject uptake at both SGI and Higher levels displays clear patterns of gender segregation, the analysis indicates that previous SGI level attainment exerts a consistent, positive impact on the choice of and attainment in STEM subjects at H level. Raising performance levels in general is, therefore, important for increasing STEM uptake at Higher. The stark significance of socio-economic background, as measured by SIMD, in both choosing named sciences at SGI level and obtaining middle or high passes, reveals the existence of a fundamental social inclusion problem in respect of STEM education in Scottish secondary schools. Given that no indicative evidence of any TTT impact was found, in line with much other research, the main reason for differences in the educational choices and attainment of boys and girls would appear to be environmental. Policy interventions to meet Scottish Government 2022 STEM uptake and attainment targets at National 5 (Maths) and Higher (Physics and Computing), need to start in schools *before* National 5 level to mitigate social exclusion and gender imbalance effects to encourage uptake at this level. Simply increasing provision/exposure at earlier secondary school levels and raising general attainment at National 5 level may not be enough; more work is needed to understand young people's aspirations and gendered choices.

Conclusion

The distribution of education in Scotland at senior secondary school and tertiary levels has been increased through policy measures that have widened access, raised participation rates and overall attainment levels. Secondary school participation beyond the age 16, the end of compulsory education, as reported earlier, has increased greatly; only some 11% of S4 pupils now leave school and more than 60% stay on until the S6, age-18, final year. Nevertheless, inequalities in educational attainment between the most disadvantaged and the most affluent persist. While the provision of tertiary education has expanded greatly, those from better-off households have benefited proportionately more. Gender inequalities are also evident with boys, in general, being seen to underachieve relative to girls since the 1970s but tending to fare better in subjects traditionally perceived as 'male': Maths and science. Socio-economic effects, however, have been found to have a greater impact on educational attainment than gender effects. The proportion of individuals from SIMD 1, the 20% most deprived households, leaving school with at least one SCQF Level 5 qualification (National 5) or above is 20% less than that for individuals from SIMD 5, the 20% least deprived households. At SCQF Level 6 qualification (Higher), it is 39% less. These inequalities have direct labour market as well as tertiary education consequences. Previous research has shown that obtaining senior school qualifications has a positive impact on individuals' earnings and that Scottish Highers attract greater labour market premia than other equivalent UK qualifications.

For policy evaluation purposes, attainment tends to be measured in terms of a benchmark number of awards at a given level (or associated number of tariff points). The weakness of this approach is that a qualification at a given level in one subject is equivalent to a qualification in any other subject at the same level. For instance, Higher Maths is regarded as equivalent to a Higher in Business Management or Physical Education. This is unsatisfactory because it ignores potentially serious inequalities related to subject choice; that is, who does what. Previous research has shown that educational inequalities in HE are greatest at the point of entry, where those from more disadvantaged households are significantly less likely to possess the necessary qualifications. Some subjects have more currency than others when applying for tertiary education courses. The prestigious Russell Group of universities advises prospective students to study certain, traditional academic subjects as these tend to facilitate entry to their institutions and courses, viz: English, Geography, History, Modern Languages, Maths, Biology, Chemistry, Physics.

The educational transitions' literature distinguishes between primary and secondary inequalities. Primary inequalities exist when differential educational attainment occurs as a direct result of socio-economic disadvantage, whereby the attainment of those from more deprived households is systematically, significantly lower than that of those from more affluent backgrounds. Secondary inequalities occur when differential, potentially less advantageous, educational choices are made by individuals from different socio-economic groups even when they have the same level of achievement. Purely focussing on the number of awards obtained at given levels by different socio-economic groups, as much policy analysis tends to do, is an evaluation of primary inequalities and ignores more nuanced secondary effects. Poor educational choices made at age-16 have the potential to impact adversely on individuals' life chances in terms of tertiary education and/or employment outcomes.

Among facilitating subjects, the uptake of Maths and the named sciences is crucial to the Scottish economy given the central importance of STEM and STEM-related industries in terms of output and jobs' growth. The rapid increase in relatively well-paid STEM jobs in Scotland is expected to continue but against a backdrop of worsening skill shortages that may act as a brake on STEM-driven economic growth and employment. Socio-economic and gender inequalities in the study of STEM subjects simultaneously reduce the future employment prospects of individuals and contribute to this labour supply bottleneck. To the extent that individuals from more deprived backgrounds are less likely to choose to study STEM subjects in secondary school, this will narrow their future employment and earnings' opportunities. STEM sector employment is highly gendered with men over-represented in architecture, engineering and construction and women over-represented in health (medicine, dentistry and general health care jobs). This gender segregation in STEM employment reflects the uptake of degree programmes at university and, before that, the uptake of STEM subjects in secondary school that directly determines HE entry and course type.

Contribution of the Thesis

The persistence of these educational inequalities makes it necessary to evaluate the allocation of education in terms of exactly *who* does *what* and *how well* they do. This thesis contributes to the literature on attainment and subject choices in Scotland by widening the lens on educational inequalities to provide more detailed insights into educational transitions, who does what and how well they do in Scottish state secondary education. It does this by exploiting national SQA administrative data for the years 2002-2009 to investigate subject choice and attainment in specific *facilitating* subjects (the academic subjects required for entry to more prestigious universities and/courses). The analyses presented are unique as, hitherto, previous work for Scotland has been survey-based (although more recently linked to matched administrative data) and focussed on the number of awards at given levels rather than the choice of and attainment in *specific* subjects. Additionally, the thesis uniquely explores whether there might be a biological determinant of the long-established gendered pattern of subject choices in schools by investigating the influence of potential exposure to heightened levels of testosterone.

The findings of the thesis provide additional insights into the nature of educational inequalities in Scottish state secondary schools in the period just before the implementation of the major CfE reform of the educational system. The thesis therefore provides both benchmarks by which to evaluate the progress made in addressing educational inequalities under the CfE and analyses of subject choice and attainment outcomes that would help to guide policy in general under the new structure. The general findings confirm much of the earlier, survey-based, empirical research on school attainment in Scotland but offer the reassurance of robustness as they are based on the analysis of data that were, in effect, a national census of achievement.

Data

SQA is the single qualifications' accreditation and awards' body for Scotland. The data, hitherto not fully exploited, were comprised of three separate, linked datasets: Candidate Results (subject, grade, level), Candidate Details (candidate number, name, age, address, postcode, urban/rural location), Centre details (centre number, type, name, address, postcode, local authority, urban/rural location). They were rationalised and standardised where necessary (e.g. multiple qualification codes, differently recorded address details).

Candidate and centre postcodes were used to link to Scottish Index of Multiple Deprivation (SIMD) information to provide the measure of socio-economic background and LFS youth employment rates were added for local authority areas. The data were rendered further to identify households and latterly twins and closelyspaced siblings. They were used to explore the determinants of subject choice and subsequent attainment using logistic regression methods to model binary subject choice decisions and the multiple alternative grade outcomes.

The data were both limited and complex to use in a number of ways. Whilst the data provide a census of attainment in Scottish state secondary schools vis à vis the outcome of individuals' entries for formal examinations or coursework, they do not contain information on those who may have embarked on the study of various qualifications and then withdrawn before being entered for assessment. If dropping out of studies is to be considered as akin to failing formal assessments then, the number of failures is under-reported and potentially gives rise to the issues of measurement error and omitted variable bias. As administrative data are not collected for research purposes, data entry can be highly variable. Data may be entered inconsistently by different individuals across years or more superfluous data, from an administrative perspective, may not be entered at all and can result in a loss of observations. The local authorities in which schools were located, for instance, were not always reported and would have resulted in the loss of 5% of the observations had they not been searched for and attached. Inconsistent entry of address information and spelling of surnames also resulted in some loss of observations when attempting to identify siblings and therefore assign household IDs. The flexibility of the Scottish Education System, allowing individuals to reach back and take lower level awards as they progress to more senior years, means that any analysis of attainment using the data must take account of the timing of awards with respect to individuals' school careers to obtain an accurate, *in-cohort* picture of achievement. The data lack the detailed background information on individuals that is to be found in large scale sample surveys for example, there is no explicit information on household income, parental occupation or individual entitlement to free school meals (which is often used a proxy for household deprivation/income). Hence the data were linked, via individuals' postcodes, to the SIMD to provide a measure of socio-economic background.

Using SIMD as an indicator of socio-economic background is potentially problematic as it is not an entirely accurate indicator of such because it is possible that individuals from low-income households may live in relatively affluent areas and *vice versa*. Entitlement to FSM is an alternative measure that has been used previously by researchers but, as discussed in the Introduction, it fails to capture many relatively poor children who just miss qualifying for such. SIMD is, in fact, a much richer variable combining information on deprivation across domains that should reflect family or neighbourhood characteristics that would be otherwise missing. Education domain indicators, for instance, capture potentially important influences on the aspirations and motivations of young people in terms of socio-cultural capital, family and neighbourhood effects that help to alleviate the potential for omitted variable bias.

Findings

Chapter One investigated the existence of secondary effects (or inequalities) in post-16 subject choice, focussing on the educational transition to the study of academic subjects at upper secondary level, the critical level of qualification for HE entry. Individuals' post-16 educational choices were examined in terms of, firstly, the decision to stay on at school to take Highers, and, secondly, the decision to study four or more Highers in facilitating subjects; this being the minimum number normally required for entry to more prestigious universities. Chapter Two then examined attainment in individual facilitating subjects in state secondary schools, estimating separate multinomial logit models for each facilitating subject at all qualification levels. Chapter Three investigated whether gendered STEM subject choice and attainment in Scottish secondary schools might have a biological component in addition to being socially constructed as generally accepted. Specifically, the potential impact of varying exposure to testosterone on STEM subject choice and attainment was explored.

Socio-economic background effects were seen to be greater than gender effects as found in previous Scottish and UK research. It was clear that individuals from the most deprived 20% of households were substantially less likely than those from the least deprived 20% of households to stay on at school after age 16 to pursue senior secondary qualifications (OR: 54%) and, if they did so, were even less likely to follow the top academic track route of taking the four or more Highers in facilitating subjects (OR: 26%). In general, to take a

Higher in a subject in fifth year, pupils had to have achieved a Credit level Standard Grade or Intermediate 2 grade C in the same subject in fourth year; SGIs (now replaced by National 5s) were *passport* qualifications to the next level. It was found that when individuals from more disadvantaged households were entered for facilitating SGIs, their likelihood of achieving the top pass necessary to proceed to the study of these subjects at Higher fell sequentially as household deprivation levels increased. In the core compulsory subjects of English and Maths, for instance, individuals from SIMD 1, compared to those from SIMD 5, were just over 50% as likely to obtain a low pass (as opposed to failing) but only 3% (English) and 6% (Maths) as likely to achieve a high pass. The analysis of STEM subject choices at SGI level, using the Twins/CSS sub-sample and, thereby controlling for family background and genetics (to the extent that this is possible), revealed further stark evidence of the influence of socio-economic background. The likelihood of taking a named science was seen to reduce, and that of taking general Science rise, sequentially with increasing disadvantage (i.e. moving from least deprived - SIMD 5 to most deprived - SIMD 1). On average, for individuals from SIMD 1, compared to those from SIMD 5, the probability of taking general Science was increased by 11% and the probability of taking other STEM subjects reduced by between 5% to 20%. In terms of attainment, being from an SIMD 1 household (as opposed to SIMD 5) in the Twins/CSS sub-sample, reduced the probability of a high pass by, on average, 18% -25% in STEM subjects at SGI level (apart from general Science, where socio-economic background was not significant). The uptake of general Science at SGI level effectively curtails the possibility of studying STEM subjects at Higher and, consequently STEM or related courses at HE (or FE) level and, thereby, access to the STEM labour market where the growth of employment opportunities is greatest.

While the well-documented, gendered nature of subject choices and attainment was evident, it was found to be more nuanced than previous Scotland- and wider UK-based research that has tended to focus on the numbers of awards at given levels. No evidence was found of any biological testosterone effect on STEM subject choices or attainment at any qualification level. Girls generally outperformed boys at age 16, being more likely to pass all Standard Grade / Intermediate 2 qualifications at all grades. There was a clear pattern of gender segregation across the STEM subjects at all qualification levels, with males more likely to take all STEM subjects apart from Biology. STEM subject choice and attainment was notably gendered at Higher, the critical level of qualification for university entry, with proportionately less females choosing STEM subjects and those that did, performing less well than males across all these subjects, at all pass grades. Males were seen to be 70% more likely than females to achieve high passes in Maths and Biology and more than twice as likely to obtain high passes in Chemistry and Physics. At Advanced Higher, male and female fortunes were reversed in Maths and science subjects, with the former being less likely to obtain any given grade in such at this level. (There was no consistent pattern of gender effects in humanities and languages facilitating subjects at either Higher or Advanced Higher.)

The only variable that was consistently significant across all the individual subject choice and attainment models at Higher was an individual's previous attainment/ability as measured by their total SGI UCAS points. The impact of this ranged from between a 1% to 4% increase in the odds of taking a given STEM Higher for a one-point increase in an individual's SGI UCAS points. This translated into a small (less than 1%), positive increase in the probability of taking a STEM Higher for one extra SGI UCAS point. Controlling for family background with the Twins/CSS models, the impact of an extra SGI UCAS point varied between a 2% increase in the likelihood of obtaining a low pass (as opposed to a fail) across all subjects and a 10% increase in the likelihood of gaining a high pass in Biology. On average, the marginal impact of an extra SGI UCAS point was to increase the probability of obtaining a middle or high pass by between 0.001% and 0.009%. SIMD effects were almost non-existent in all the Twins/CSS Higher models; where they did occur, there was no consistent pattern. The Twins/CSS' results would tend to suggest that, in keeping with the Behavioural Genetics' literature, once family background is controlled for in an egalitarian educational system, much of the remaining variation in children's educational attainment is at the level of the individual.

Policy Relevance and Implications for Further Research

There are clear implications from the findings for the operation of the CfE and education policy. Subject choices made at Higher are crucial as these determine HE entry, course type and future occupation, these choices are fed by age 14 subject choices and attainment therein at age 16. The structure of the curriculum in Scotland underpinning age-16 qualifications during the years examined, ensured that a minimum number of academic subjects were studied, providing a broad-based education covering the core disciplines of English and Maths and then a science, humanity and modern language. Where there was choice at this level, individuals from the most deprived 40% of households were seen to be less likely to choose to study, additional academic subjects; for instance, a second science or modern language. Given the propensity of those from more deprived backgrounds to not choose named sciences at SGI level and the reduced likelihood of girls taking Physics in particular, the increased individual choice embedded in the CfE could inadvertently worsen these horizontal inequalities. Add to this the evidence to suggest that the number of subjects now studied for formal qualifications (National 5s) at age 16 can be as few as five in some schools (as opposed to seven or eight commonly under the SGI structure), increased individual choice may be occurring against a background of reduced options. This in turn may further exacerbate inequalities of subject choice and, indeed, there is some evidence to suggest that the uptake of sciences at 14-16 has fallen, albeit slightly, under the CfE. To assess progress in addressing educational inequality at school level thoroughly, subject choice needs be included and monitored for policy evaluation purposes and not just the bald number of awards at a given level. Given that those from more disadvantaged households tend to have access to less social capital, there is a clear need for improved information and guidance with respect to subject choices in secondary schools.

In the UK and further afield, research has shown family characteristics to dominate financial constraints and expectations of successful completion of courses to be important in making educational decisions. In Germany, the perceived probability of success has been seen to be determined by children's grade points and type of school recommended at the end of the primary stage. In Scotland, recent work using survey data has shown secondary inequalities in subject choices to explain differential HE entry. Given that children from disadvantaged backgrounds tend to perform less well at school, it is possible that child and parental educational expectations, formed on the basis of this lower achievement, may feed through into poorer, less academic, secondary school subject choices thereby producing and sustaining inequalities in HE entry over time. It is likely also that children from more disadvantaged households will have less social capital to draw on in terms of immediate family members with knowledge of how the later levels of the education system operate and, if so, this might compound poor subject choice decisions.

The gender effects for STEM subjects were far outweighed by the socio-economic effects. The stark significance of socio-economic background, as measured by SIMD, in both choosing named sciences at SGI level and obtaining middle or high passes, revealed the existence of a fundamental social inclusion problem in respect of STEM education in Scottish secondary schools. This again indicates the need to improve information and guidance around subject choice but also suggests that policy interventions to mitigate social exclusion and gender imbalance in the uptake of STEM subjects need to start in schools, building substantial science social capital long *before* subject choices for initial formal qualifications are made at age 14. Otherwise, the Scottish Government's STEM Education and Training Strategy targets of a reduction in the STEM attainment gap (to 25 percentage points) at Higher between school leavers from the least deprived and most deprived households and increases in the numbers of females passing Higher Physics (15%) and Computing (20%) by 2022 are unlikely to be realised.

Individuals' ability was clearly important in terms of securing a low pass in facilitating subjects at Higher but was seen not to be enough to overcome disadvantage to achieve the higher grades that may be required for entry to more prestigious HE institutions and/or courses. This would suggest that, once *good* subject choices have been made, there is a case for HE access thresholds, introduced in 2019, whereby prospective students from more deprived backgrounds are to be assessed against the minimum academic standards and subject knowledge needed for successful completion of a degree programme. The absence of any innate biological testosterone effect to explain marked gendered patterns of STEM subject choice and attainment suggests that, in line with much other research and as is commonly accepted, the main reason for differences in female and male educational choices would appear to be environmental.

As previous research has suggested, more qualitative work is needed to understand young people's aspirations and gendered choices. The exploitation of education administrative data, however, such as the SQA database of candidate results used here, has the potential to provide a more complete picture of attainment and answer fundamental policy questions more robustly despite lacking detailed background information on individuals. Researchers will need to turn increasingly to the use of administrative data as the funding for large-scale, detailed surveys continues to come under increasing pressure that may either lead to their curtailment or reduce their coverage and/or reliability. The increased power of computing hardware and statistical software has provided researchers with the ability to analyse enormous administrative datasets and if administrative data from different agency sources can be fully linked, this would provide richer individual information.

Appendix A

Appendix to Chapter One

Table A1

Sample Summary Statistics: Staying on to take at least one Higher Decision Model

Total Observations	**Total No Cases**	**Observed once**	**Observed twice**	**Observed 3/+**	
1,021,470	489,468	166,136	114,662	208,670	
Dependent Variable	**N: Y = 1**	**N: Y = 0**	**% Y = 1**	**% Y = 0**	
Staying on/1+ Higher	204,802	192,391	52%	48%	
Independent Variables of Main Interest	**N**		**%**		
Sex (Male = 1, Female = 0)	M: 247,210	F: 242,258	M: 50.5%	F: 49.5%	
SIMD of Household					
SIMD 1	102,211		21%		
SIMD 2	97,291		20%		
SIMD 3	95,861		20%		
SIMD 4	95,457		20%		
SIMD 5	90,073		19%		
Total SIMD Nos	480,893		2% Missing (8,575)		
Continuous Variables	**Mean**	**Standard Deviation**	**Maximum**	**Minimum**	**Obs**
SGI points	159	83	422	0	444,668
SGI Cohort Size	181	71	536	1	489,452
School % SIMD 1 & 2 pupils	39	25	100	0	489,468
LA Youth employment rate	63	8	95	43	489,468
Other Variables	**N**		**%**		
Urban/Rural Location of School attended					
Accessible Rural	55,186		11%		
Accessible Small Towns	43,556		9%		
Large Urban Areas	168,506		35%		
Other Urban Areas	161,031		34%		
Remote Rural	32,371		7%		
Remote Small Towns	20,243		4%		
Total	480,893		2% Missing (8,575)		
Local Authority of School attended					
Aberdeen City	15,743		3.5%		
Aberdeenshire	24,859		5.5%		
Angus	10,053		2%		
Argyll & Bute	8,497		2%		
Clackmannanshire	2,941		0.7%		
Dumfries & Galloway	15,743		3.5%		
Dundee City	10,779		2%		
East Ayrshire	12,150		3%		
East Dunbartonshire	11,102		2.5%		
East Lothian	8,802		2%		
East Renfrewshire	11,084		2.5%		
Edinburgh, City of	29,557		6.5%		
Eilean Siar	2,906		0.6%		
Falkirk	14,649		3%		
Fife	35,440		8%		

Table A1 cont.

Sample Summary Statistics: Staying on to take at least one Higher Decision Model

Local Authority of School attended		
Glasgow, City of	42,929	9.5%
Highland	19,764	4%
Inverclyde	2,108	0.5%
Midlothian	9,453	2%
Moray	9,315	2%
North Ayrshire	12,371	3%
North Lanarkshire	31,575	7%
Orkney Islands	2,235	0.5%
Perth & Kinross	12,122	3%
Renfrewshire	16,068	3.5%
Scottish Borders	9,431	2%
Shetland Islands	2,572	0.6%
South Ayrshire	9,227	2%
South Lanarkshire	25,521	6%
Stirling	7,787	2%
West Dunbartonshire	9,588	2%
West Lothian	16,258	3.5%
Total	452,629	7% Missing (36,776)
Year dummy		
2002	59,083	12%
2003	60,211	12%
2004	60,465	12%
2005	60,817	12%
2006	64,811	13%
2007	64,066	13%
2008	62,124	13%
2009	57,891	12%

Table A2

Sample Summary Statistics: Taking Four or more Facilitating Highers Decision Model

Total Observations:	**Total No Cases**		**Observed once**		**Observed twice**		**Observed 3/+**
1,021,470	489,468		166,136		114,662		208,670
Staying on/take 1+ Highers	204,802		204,802		204,802		
Dependent Variable	**N: Y = 1**		**N: Y = 0**		**% Y = 1**		**% Y = 0**
Taking 4+ Facilitating Highers	57,450		147,378		28%		72%
Independent Variables of Main Interest	**N**				**%**		
Sex (Male = 1, Female = 0)	M: 95,068		F: 109,760		M: 46%		F: 54%
SIMD of Household							
SIMD 1	23,769				12%		
SIMD 2	31,723				16%		
SIMD 3	40,245				20%		
SIMD 4	48,912				24%		
SIMD 5	57,165				28%		
Total SIMD Nos	201,814				1.5% Missing (2,988)		
Continuous Variables	**Mean**	**Standard Deviation**		**Maximum**		**Minimum**	**Obs**
SGI points	228	51		422		0	204,414
Higher Cohort Size	96	43		247		1	204,828
School % SIMD 1 & 2 pupils	33	23		100		0	204,828
Other Variables	**N**				**%**		
Urban/Rural Location of School attended							
Accessible Rural	25,753				13%		
Accessible Small Towns	19,46				10%		
Large Urban Areas	68,457				34%		
Other Urban Areas	63,824				32%		
Remote Rural	16,007				8%		
Remote Small Towns	8,305				4%		
Total	201,814				1.5% Missing (2,988)		
Local Authority of School attended							
Aberdeen City	6,682				3.5%		
Aberdeenshire	11,102				5.8%		
Angus	4,519				2.4%		
Argyll & Bute	4,127				2.2%		
Clackmannanshire	880				0.5%		
Dumfries & Galloway	6,776				3.6%		
Dundee City	3,667				1.9%		
East Ayrshire	5,327				2.8%		
East Dunbartonshire	6,831				3.6%		
East Lothian	3,913				2.1%		
East Renfrewshire	7,204				3.8%		
Edinburgh, City of	12,034				6.4%		
Eilean Siar	1,675				0.9%		
Falkirk	4,736				2.5%		
Fife	14,451				7.7%		

Table A2 cont.

Sample Summary Statistics: Taking Four or more Facilitating Highers Decision Model

Local Authority of School attended		
Glasgow, City of	14,382	7.6%
Highland	8,848	4.7%
Inverclyde	790	0.4%
Midlothian	2,759	1.5%
Moray	3,715	2.0%
North Ayrshire	4,559	2.4%
North Lanarkshire	11,780	6.2%
Orkney Islands	1,187	0.6%
Perth & Kinross	5,017	2.7%
Renfrewshire	7,901	4.2%
Scottish Borders	4,719	2.5%
Shetland Islands	1,397	0.7%
South Ayrshire	4,163	2.2%
South Lanarkshire	11,571	6.1%
Stirling	3,194	1.7%
West Dunbartonshire	3,707	2.0%
West Lothian	5,059	2.7%
Total	188,672	8% Missing (16,130)
Year dummy		
2002	28,263	13.8%
2003	27,872	13.6%
2004	27,542	13.5%
2005	26,828	13.1%
2006	24,933	12.2%
2007	24,758	12.1%
2008	23,003	11.2%
2009	21,629	10.6%

Table A3

Summary of Likelihood Ratio Tests / Post Estimation Tests of Coefficients

Models	**Log likelihood**	**Pseudo R2**	**Likelihood Ratio Test**	**Post estimation test of significance of βs** (unrestricted model)	**Comments**
Hstay – unrestricted [Decision to stay on to take Highers]	-86194.4	.541	203152.83***		-187770.84 Null model log likelihood
Hstay - restricted					
No SGIpts4	-100161.84	.47	27934.83***		
No Sex	-86484.764	.539	580.68***		
No SIMD2009 quintiles	-89177.648	.536		1654.94***	Inclusion -> 5,581 observations dropped (282706 without – 277125 with) -192019.02 Null model log likelihood
No SGIpts4 School quartiles	-87917.027	.532	3445.20***		
No sgicoh4 – SGI cohort size	-86208.908	.541	28.97***		

Table A3 cont.

Models	Log likelihood	Pseudo R2	Likelihood Ratio Test	Post estimation test of significance of βs (unrestricted model)	Comments
Hstay – restricted cont.					
No SIMD12percent - % pupils from most disadvantaged 40%	-86290.543	.54	192.24***		
No School urban/rural location dummies	-93601.524	.543		49.17***	Inclusion -> 24,806 observations dropped (301,931 without – 277125 with) -204664.53 Null model log likelihood
No Local Authority dummies	-87018.612	.54	1648.37***		
No year dummies	-132459.24	.41		96.35***	-> 46,584 observations dropped (323710 without – 277125 with) -224371.16 Null model log likelihood
No Youth Employment Rate	-135519.46	.396		27.98***	-> 46,584 observations dropped (323710 without – 277125 with) -224371.16 Null model log likelihood

Table A3 cont.

Models	Log likelihood	Pseudo R2	Likelihood Ratio Test	Post estimation test of significance of βs (unrestricted model)	Comments
Hfs4 – unrestricted [Decision to take Four or more Facilitating Subject Highers]	-67545.323	.388	85481.31***		
Hfs4 - restricted					
No SGIpts4	-89443.375	.189	43796.10***		
No Sex	-67556.021	.387	21.39***		
No SIMD2009 quintiles	-68580.407	.386		185.44***	Inclusion -> 2,579 observations dropped (188277 without – 185698 with)
No SGIpts4 School quartiles	-67727.132	.386	363.62***		
No hcoh5 – H cohort size	-67557.454	.387	24.26***		
No SIMD12percent - % pupils from most disadvantaged 40%	-67590.932	.387	91.22***		

Table A3 cont.

Models	Log likelihood	Pseudo R2	Likelihood Ratio Test	Post estimation test of significance of βs (unrestricted model)	Comments
Hfs4 – restricted cont.					
No School urban/rural location dummies	-73315.136	.387		46.15***	Inclusion -> 15,726 observations dropped (201424 without – 185698 with)
No Local Authority dummies	-68244.064	.381	1397.48***		
No year dummies	-67698.761	.386	306.87***		
No Youth Employment Rate	-67545.4	.388	0.15		

Table A4
Sequential Logit Analysis

	(1) **Decision to Stay on to take at least One Higher**	(2) **Decision to take Four or more Facilitating Highers**
SGIpts4	1.034***	1.045***
Sex (M=1)	0.791***	1.053
SIMD2009quintile 1	0.540***	0.740***
SIMD2009quintile 2	0.638***	0.822***
SIMD2009quintile 3	0.686***	0.860***
SIMD2009quintile 4	0.868*	0.923**
Sex SIMD 1	1.082	0.937
SIMD 2	1.008	0.922
SIMD 3	1.000	1.021
SIMD 4	0.912	1.002
SGIpts4schqrt=1	9.796***	3.907
SGIpts4schqrt=2	0.610***	0.899
SGIpts4schqrt=3	0.806**	0.684***
Sex SGIpts4schqrt 1	0.658***	3.394
SGIpts4schqrt 2	0.800***	0.928
SGIpts4schqrt 3	0.971	1.063
SGIpts4schqrt 1 SIMD 1	0.147***	14.288*
SIMD 2	0.184***	19.472
SIMD 3	0.372***	0.316
SIMD 4	0.443***	1.084
SGIpts4schqrt 2 SIMD 1	0.518***	3.120
SIMD 2	0.502***	3.086**
SIMD 3	0.682***	0.910
SIMD 4	0.719***	0.771
SGIpts4schqrt 3 SIMD1	0.757***	0.803
SIMD 2	0.818**	0.812*
SIMD 3	0.895	0.781**
SIMD 4	0.861**	0.893
sgicoh4	0.999	1.001
SIMD12percent	0.994***	0.994*
Accessible Rural	0.974	1.137
Accessible Small Towns	0.901	0.922
Other Urban Areas	0.963	0.998
Remote Rural	1.129	1.055
Remote Small Towns	1.105	1.118
Aberdeen City	0.820	0.807
Aberdeenshire	0.432***	0.403***
Angus	0.742*	0.836
Argyll & Bute	0.813	0.594*
Clackmannanshire	0.688***	0.616*
Dumfries & Galloway	0.751	0.495**
Dundee City	0.978	0.666**

Table A4 cont.
Sequential Logit Analysis

	(1) **Decision to Stay on to take at least One Higher**	(2) **Decision to take Four or more Facilitating Highers**
East Ayrshire	0.831	0.683*
East Dunbartonshire	0.952	0.810
East Lothian	0.515**	0.459**
East Renfrewshire	2.265***	0.511***
Edinburgh, City of	0.807	0.523***
Eilean Siar	2.088***	1.634**
Falkirk	0.981	0.593**
Fife	0.832	0.512***
Highland	0.955	0.599**
Inverclyde	1.052	1.274
Midlothian	0.635	0.472***
Moray	0.468**	0.272***
North Ayrshire	0.924	1.189
North Lanarkshire	0.925	0.880
Orkney Islands	0.779	0.692
Perth & Kinross	0.743	0.543**
Renfrewshire	0.838	0.793
Scottish Borders	0.799	0.472***
Shetland Islands	0.411***	0.497***
South Ayrshire	0.885	0.624*
South Lanarkshire	0.931	0.829
Stirling	1.183	0.738
West Dunbartonshire	1.124	0.739*
West Lothian	0.764	0.468***
2002	0.887*	1.390***
2003	0.872*	1.279***
2004	0.858**	1.210***
2005	0.833**	1.096
2006	0.874*	0.990
2007	0.982	0.989
2008	1.000	1.114**
Female/Male Youth Employment	0.992***	-
Observations	277125	185698
Log pseudolikelihood	-86194.425	-67545.4
Wald chi2(73)	25021.71***	15662.55***
Pseudo R2	0.541	0.388

Notes:
Exponentiated coefficients (odds ratios) give the change in the odds associated with a one unit increase in a given independent variable holding other variables constant. SIMD 1-5 = Scottish Index of Multiple Deprivation quintiles
Omitted reference categories: SIMD 5 – the least deprived (most affluent) 20% of households, SGIpts4schqrt 4 – top school SGI UCAS points' quartile
Other independent variables included: Urban/Rural location, Local Authority and Year fixed effects
Cluster-robust standard errors at the school level (Model 1: 352, Model 2: 348)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5

Sequential Logit Analysis: Local Authority Odds Ratios - descending order of magnitude

Local Authority	Staying on to take at least one Higher		Local Authority	Taking Four or more Facilitating Highers	
East Renfrewshire	2.265	***	Eilean Siar	1.634	**
Eilean Siar	2.088	***	Inverclyde	1.274	
Stirling	1.183		North Ayrshire	1.189	
West Dunbartonshire	1.124		North Lanarkshire	0.88	
Inverclyde	1.052		Angus	0.836	
Falkirk	0.981		South Lanarkshire	0.829	
Dundee City	0.978		East Dunbartonshire	0.81	
Highland	0.955		Aberdeen City	0.807	
East Dunbartonshire	0.952		Renfrewshire	0.793	
South Lanarkshire	0.931		West Dunbartonshire	0.739	*
North Lanarkshire	0.925		Stirling	0.738	
North Ayrshire	0.924		Orkney Islands	0.692	
South Ayrshire	0.885		East Ayrshire	0.683	*
Renfrewshire	0.838		Dundee City	0.666	**
Fife	0.832		South Ayrshire	0.624	*
East Ayrshire	0.831		Clackmannanshire	0.616	*
Aberdeen City	0.82		Highland	0.599	**
Argyll & Bute	0.813		Argyll & Bute	0.594	*
Edinburgh, City of	0.807		Falkirk	0.593	**
Scottish Borders	0.799		Perth & Kinross	0.543	**
Orkney Islands	0.779		Edinburgh, City of	0.523	***
West Lothian	0.764		Fife	0.512	***
Dumfries & Galloway	0.751		East Renfrewshire	0.511	***
Perth & Kinross	0.743		Shetland Islands	0.497	***
Angus	0.742	*	Dumfries & Galloway	0.495	**
Clackmannanshire	0.688	***	Scottish Borders	0.472	***
Midlothian	0.635		Midlothian	0.472	***
East Lothian	0.515	**	West Lothian	0.468	***
Moray	0.468	**	East Lothian	0.459	**
Aberdeenshire	0.432	***	Aberdeenshire	0.403	***
Shetland Islands	0.411	***	Moray	0.272	***

Table A6

Local Authority Areas: Correlation between Decision to Stay to take at Least One Higher & Decision to take Four or more Facilitating Highers

Local Authority	Staying on to take at least one Higher		Rank	Local Authority	Taking Four or more Facilitating Highers		Rank
Aberdeen City	0.82		17	Aberdeen City	0.807		8
Aberdeenshire	0.432	***	30	Aberdeenshire	0.403	***	30
Angus	0.742	*	25	Angus	0.836		5
Argyll & Bute	0.813		18	Argyll & Bute	0.594	*	18
Clackmannanshire	0.688	***	26	Clackmannanshire	0.616	*	16
Dumfries & Galloway	0.751		23	Dumfries & Galloway	0.495	**	25
Dundee City	0.978		7	Dundee City	0.666	**	14
East Ayrshire	0.831		16	East Ayrshire	0.683	*	13
East Dunbartonshire	0.952		9	East Dunbartonshire	0.81		7
East Lothian	0.515	**	28	East Lothian	0.459	**	29
East Renfrewshire	2.265	***	1	East Renfrewshire	0.511	***	23
Edinburgh, City of	0.807		19	Edinburgh, City of	0.523	***	21
Eilean Siar	2.088	***	2	Eilean Siar	1.634	**	1
Falkirk	0.981		6	Falkirk	0.593	**	19
Fife	0.832		15	Fife	0.512	***	22
Highland	0.955		8	Highland	0.599	**	17
Inverclyde	1.052		5	Inverclyde	1.274		2
Midlothian	0.635		27	Midlothian	0.472	***	26
Moray	0.468	**	29	Moray	0.272	***	31
North Ayrshire	0.924		12	North Ayrshire	1.189		3
North Lanarkshire	0.925		11	North Lanarkshire	0.88		4
Orkney Islands	0.779		21	Orkney Islands	0.692		12
Perth & Kinross	0.743		24	Perth & Kinross	0.543	**	20
Renfrewshire	0.838		14	Renfrewshire	0.793		9
Scottish Borders	0.799		20	Scottish Borders	0.472	***	26
Shetland Islands	0.411	***	31	Shetland Islands	0.497	***	24
South Ayrshire	0.885		13	South Ayrshire	0.624	*	15
South Lanarkshire	0.931		10	South Lanarkshire	0.829		6
Stirling	1.183		3	Stirling	0.738		11
West Dunbartonshire	1.124		4	West Dunbartonshire	0.739	*	10
West Lothian	0.764		22	West Lothian	0.468	***	28

Rank Correlation: 61%

Figure A1: Decision to stay on & take Highers :
Scatterplot and local mean regression of Standard Grade/Intermediate 2 points

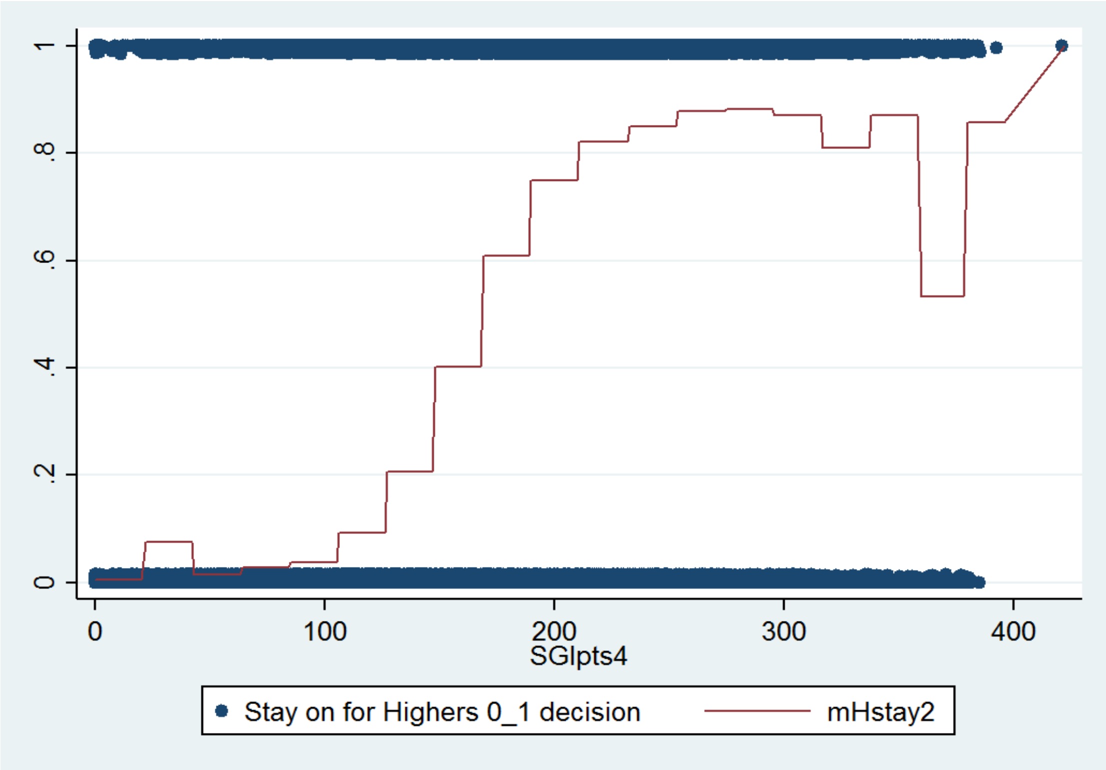


Figure A2: Decision to stay on & take Highers :
Scatterplot and local mean regression of Standard Grade/Intermediate 2 cohort size

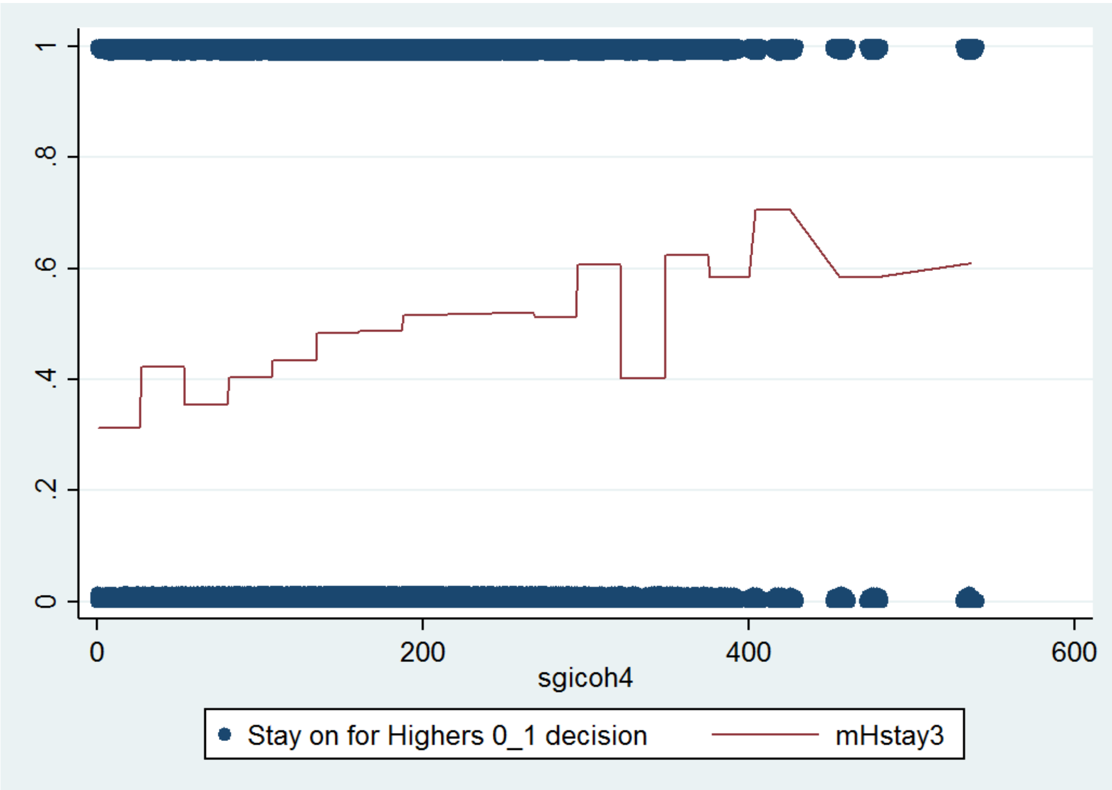


Figure A3: Decision to stay on & take Highers :
Scatterplot and local mean regression of school percentage of SIMD 1 & 2 pupils

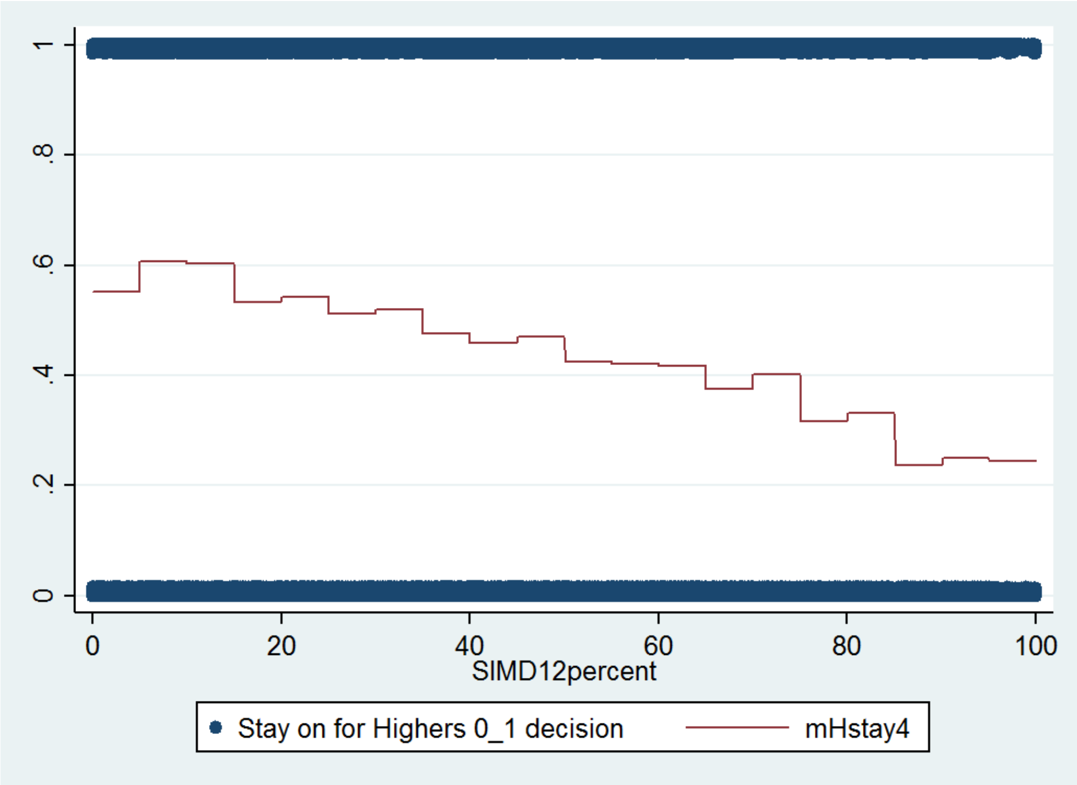


Figure A4: Decision to stay on & take Highers:
Scatterplot and local mean regression of LA youth employment rate

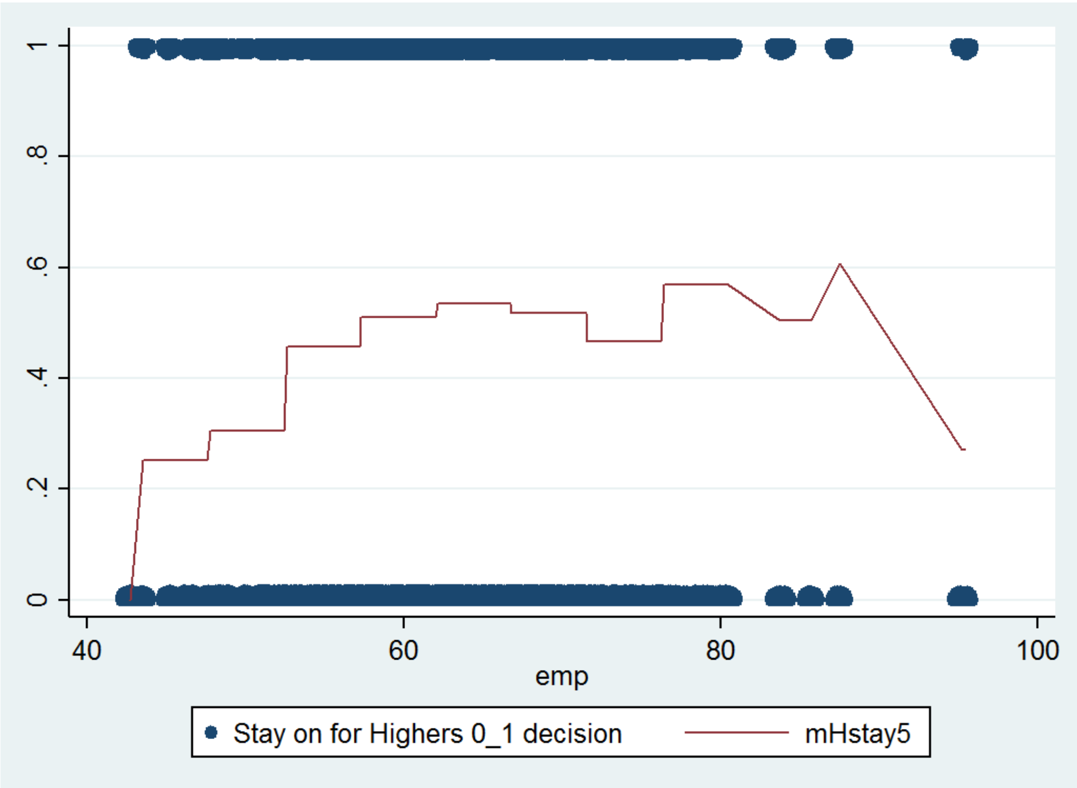


Figure A5: Decision to stay on & take Highers: Discrepancy v Leverage

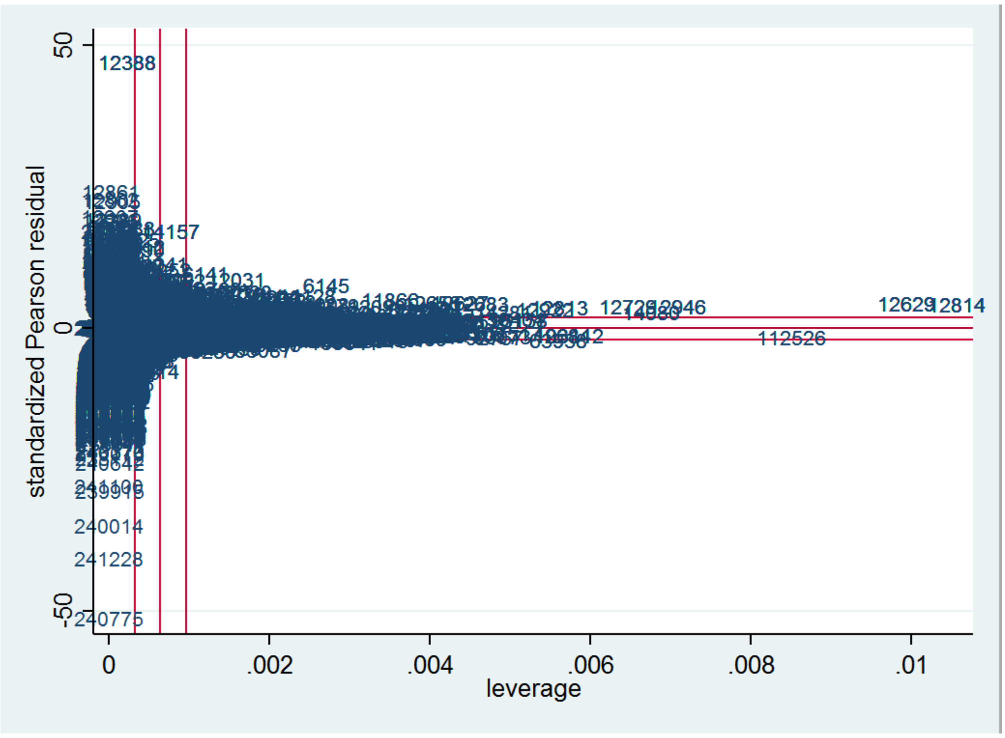
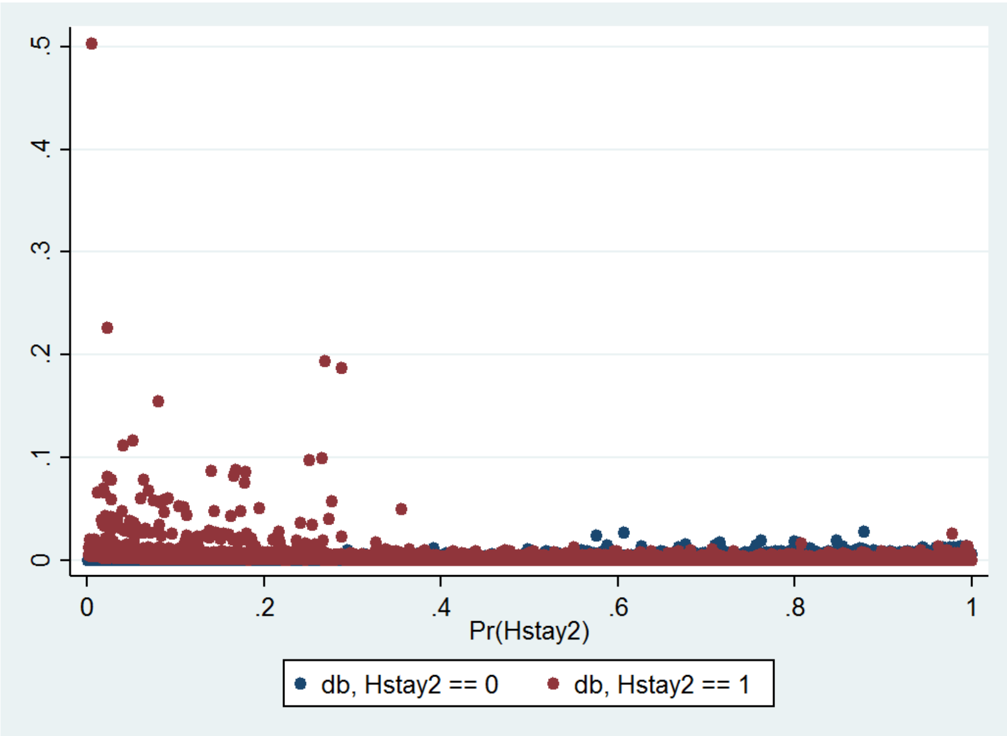


Figure A6: Decision to stay on & take Highers: Δβ v Predicted Probabilities [96]



[96] Δβ is the equivalent for logistic regression of the Cook's D statistic for linear regression and estimates the effect of one covariate pattern on all regression coefficients simultaneously.

Figure A7: Decision to stay on & take Highers: $\Delta\chi^2_{P(j)}$ v Predicted Probabilities [97]

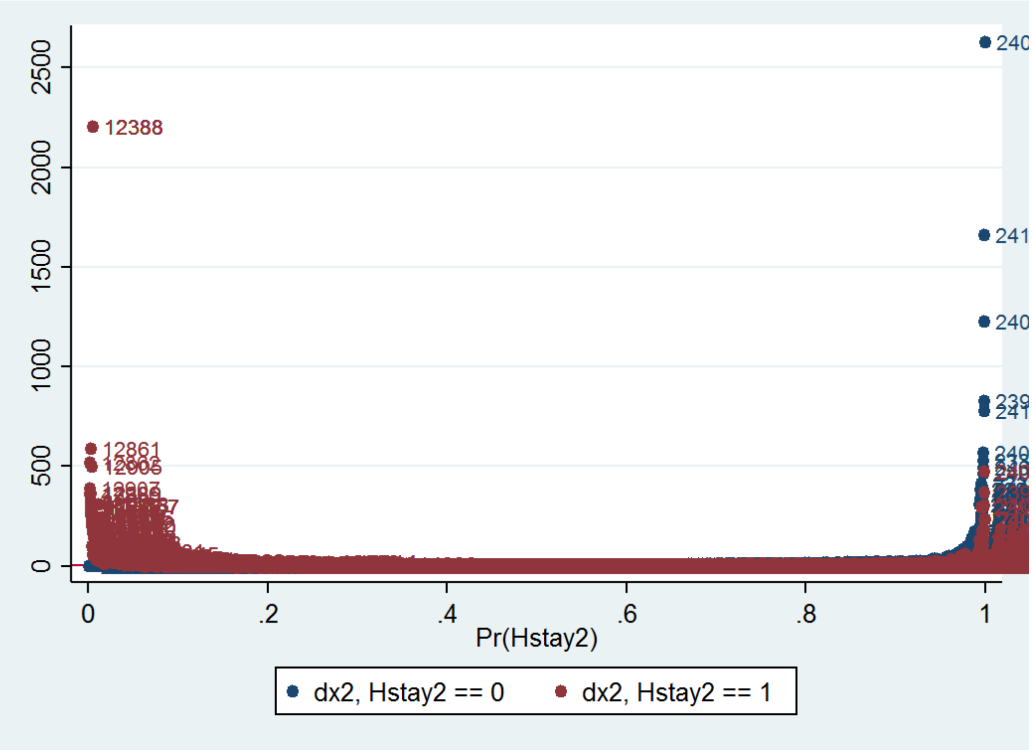
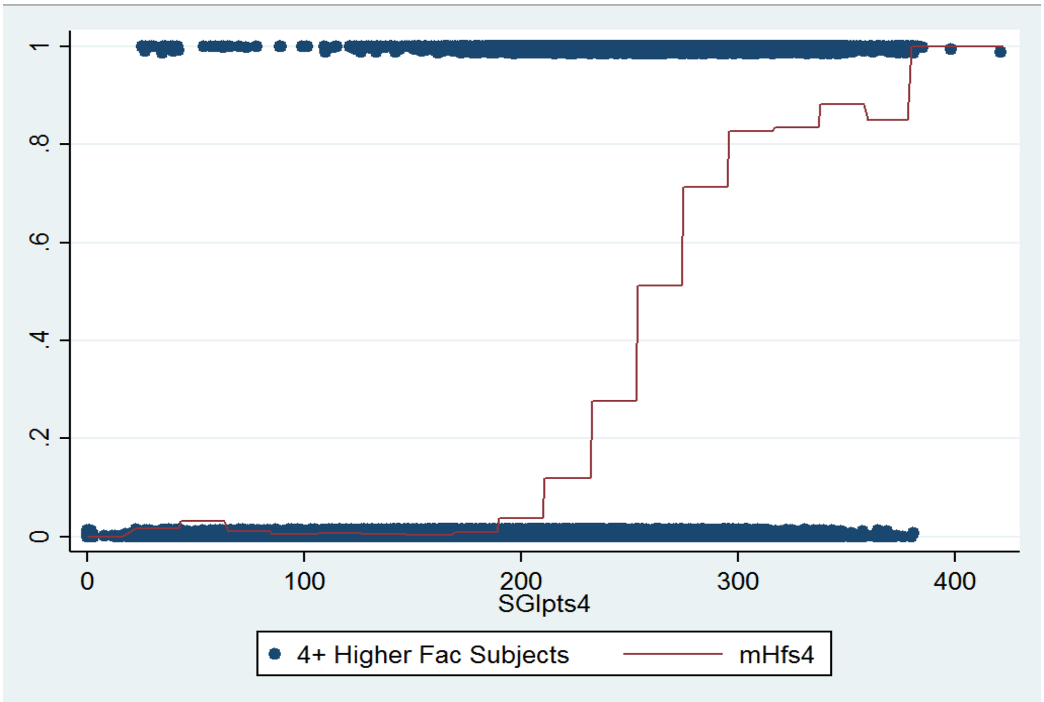


Figure A8: Decision to take Four or more Facilitating Highers: Scatterplot and local mean regression of Standard Grade/Intermediate 2 points



[97] Change in Pearson Chi-sq when covariate pattern *j* is removed from dataset.

Figure A9 Decision to take Four or more Facilitating Highers: Scatterplot and local mean regression of Higher cohort size

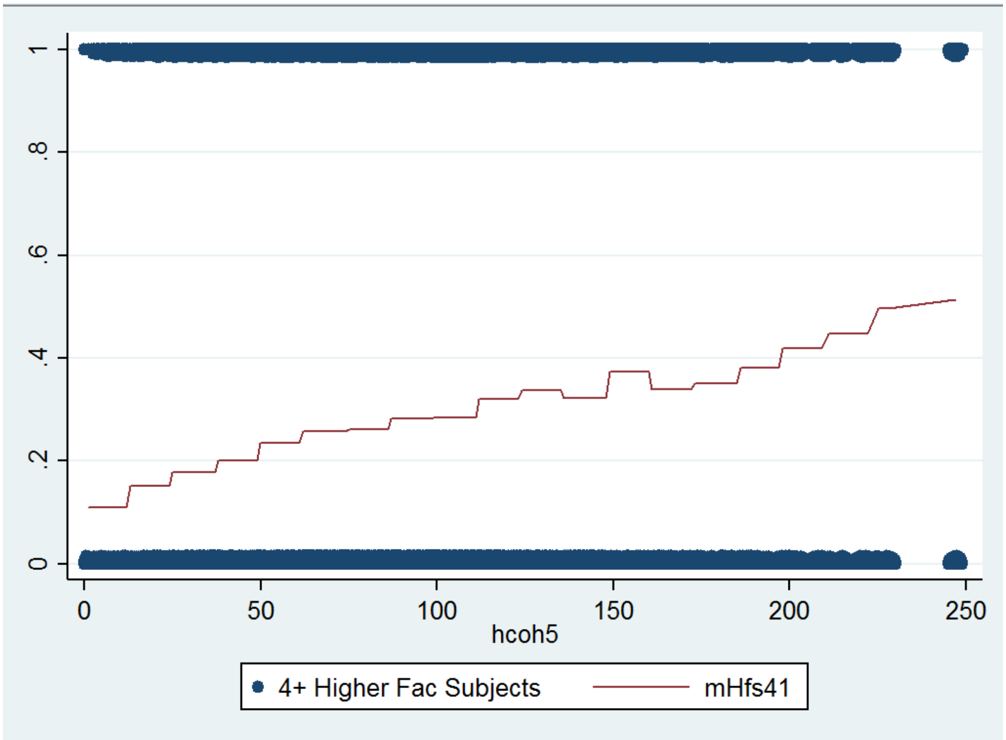


Figure A10: Decision to take Four or more Facilitating Highers: Scatterplot and local mean regression of school percentage of SIMD 1 & 2 pupils

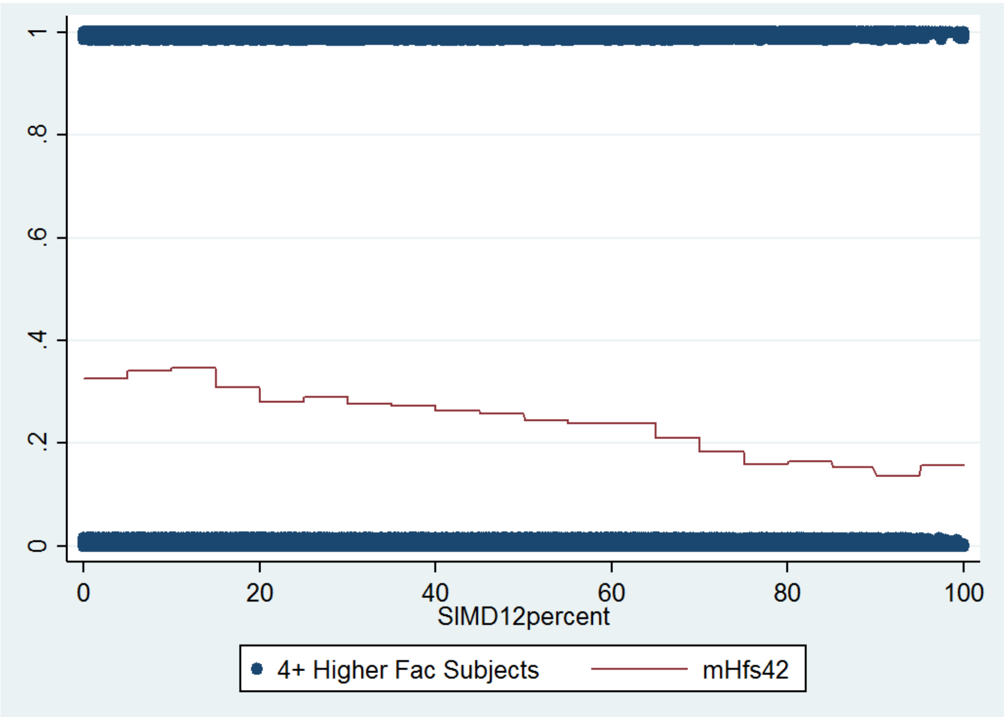


**Figure A11: Decision to take Four or more Facilitating Highers:
Scatterplot and local mean regression of LA youth employment rate**

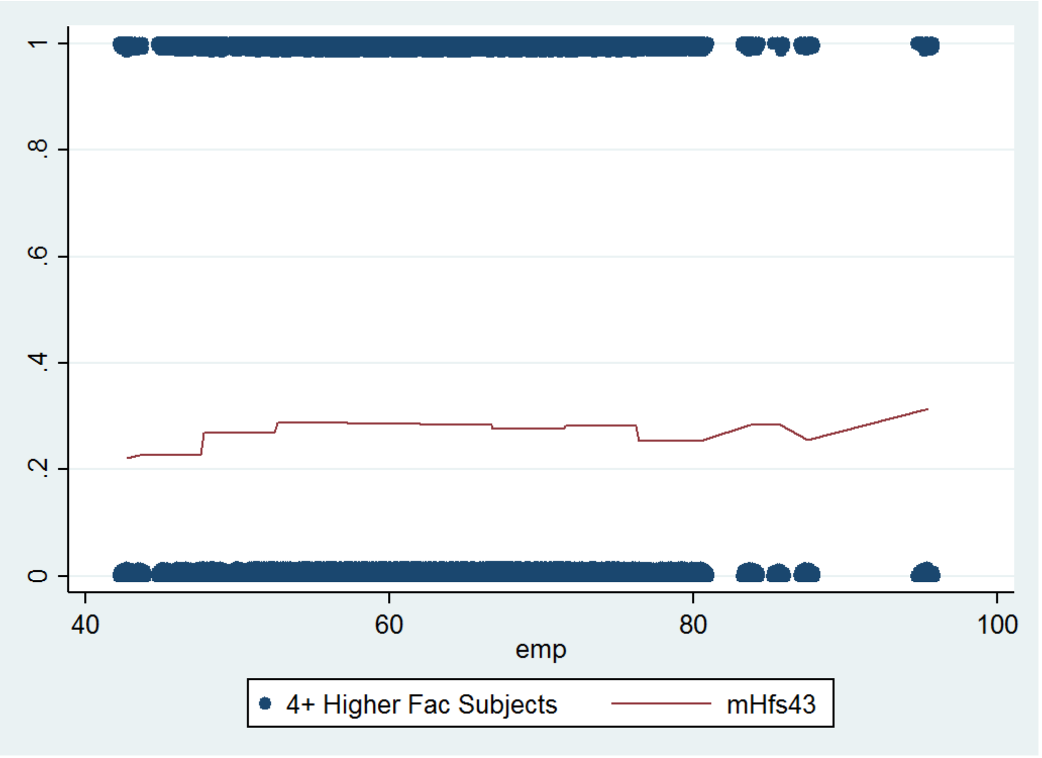


Figure A12: Decision to take Four or more Facilitating Highers: Discrepancy v Leverage

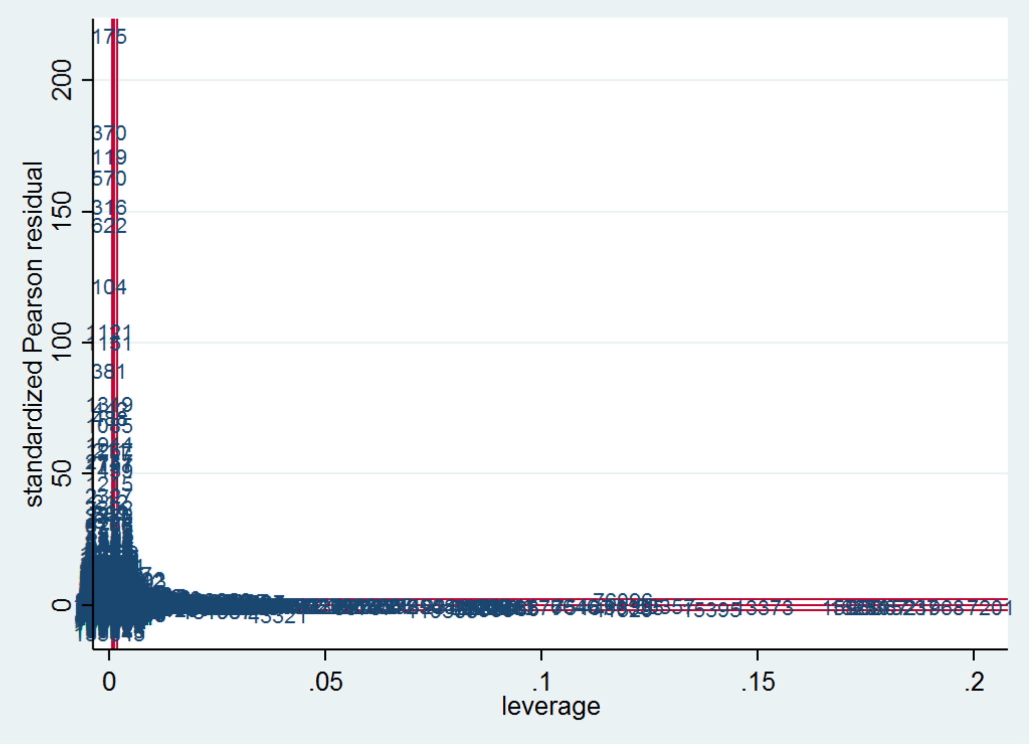


Figure A13: Decision to take Four or more Facilitating Highers: $\Delta\beta$ v Predicted Probabilities

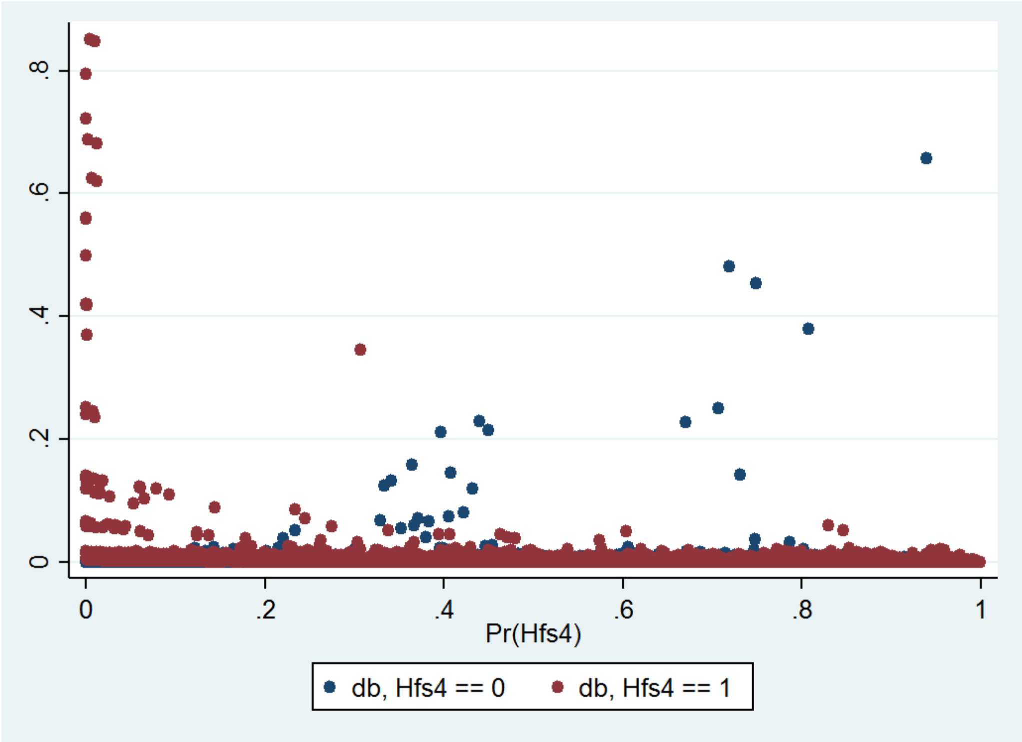


Figure A14: Decision to take Four or more Facilitating Highers: $\Delta\chi^2_{P(j)}$ v Predicted Probabilities

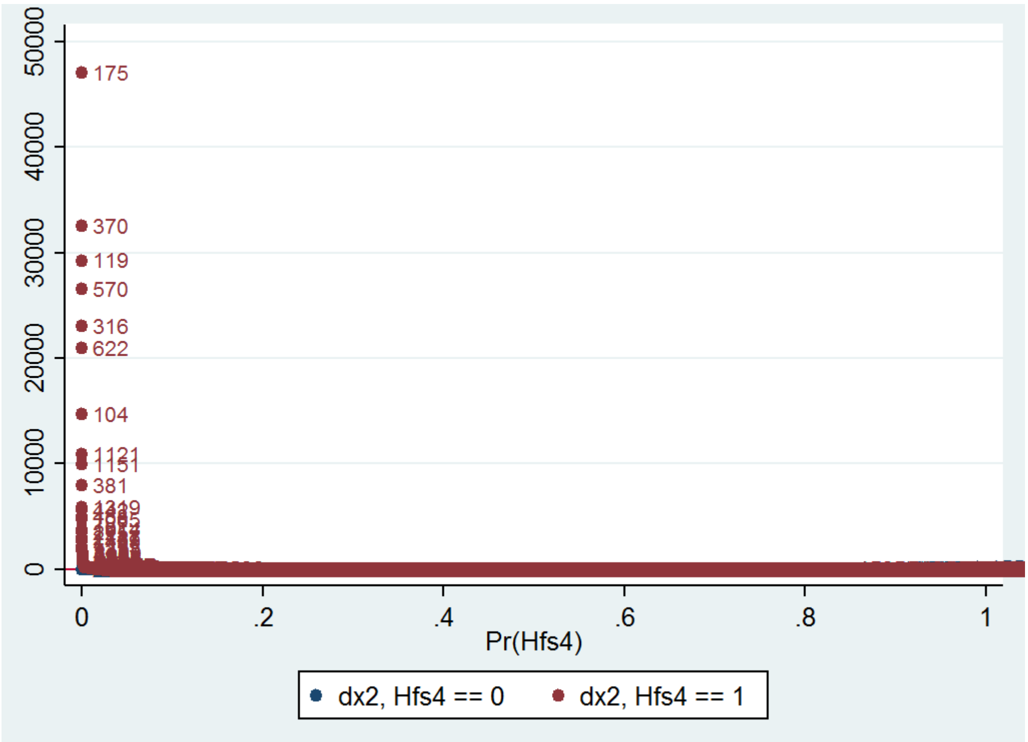
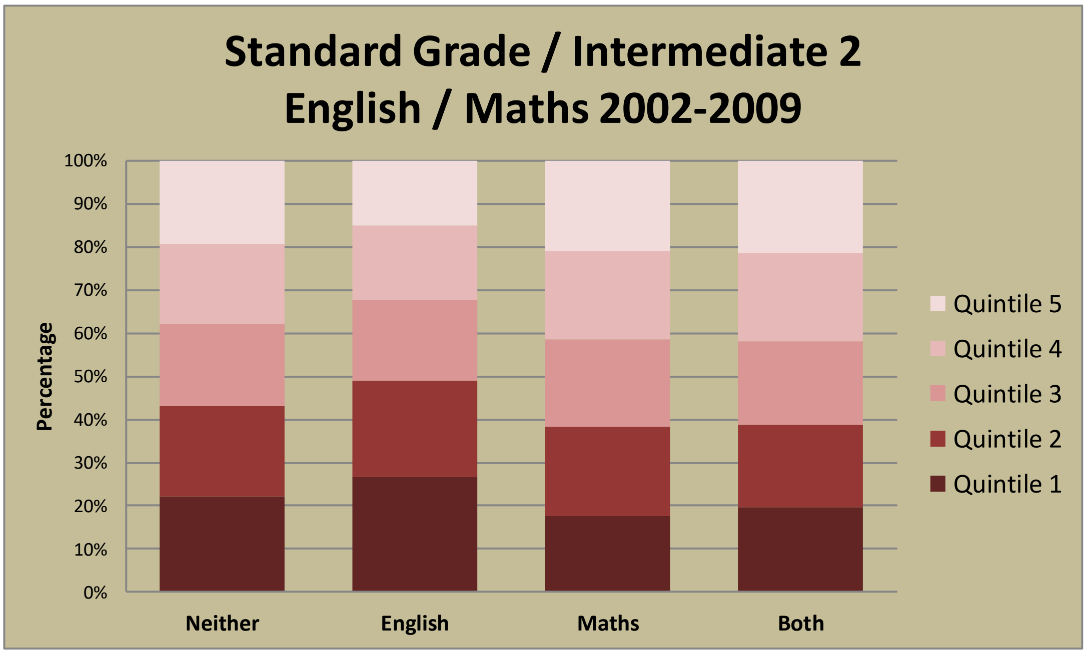
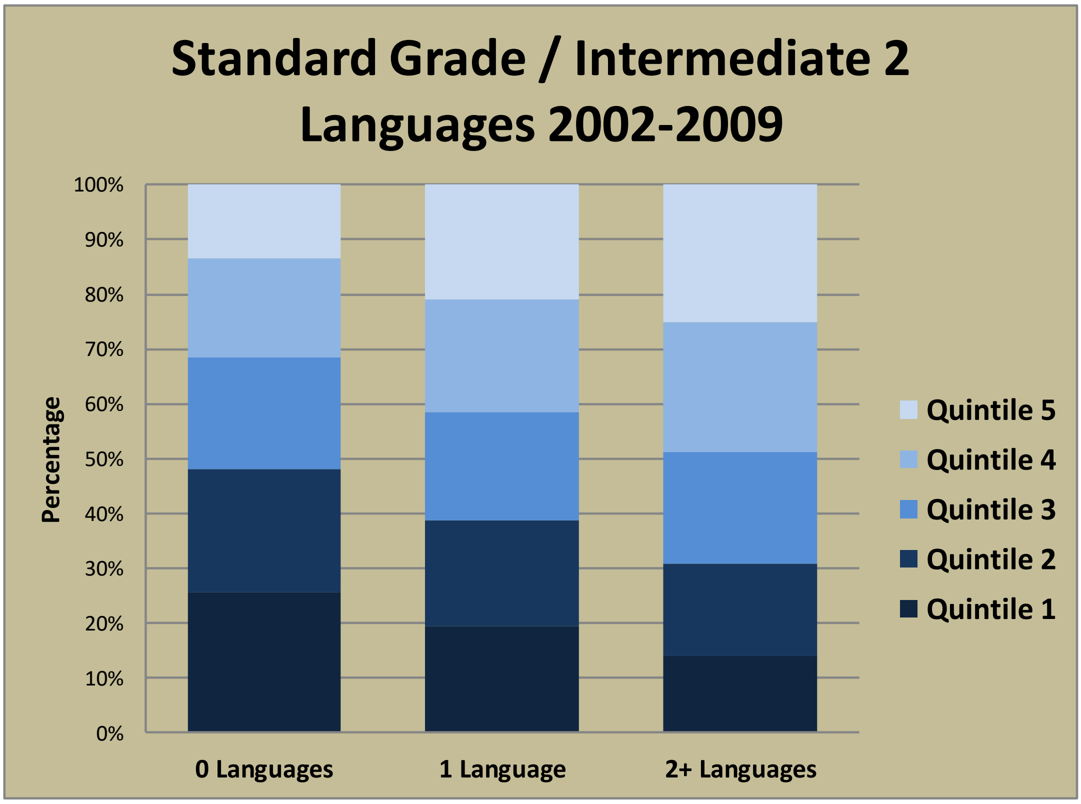


Figure A15: English and Maths Studied in S4 by SIMD



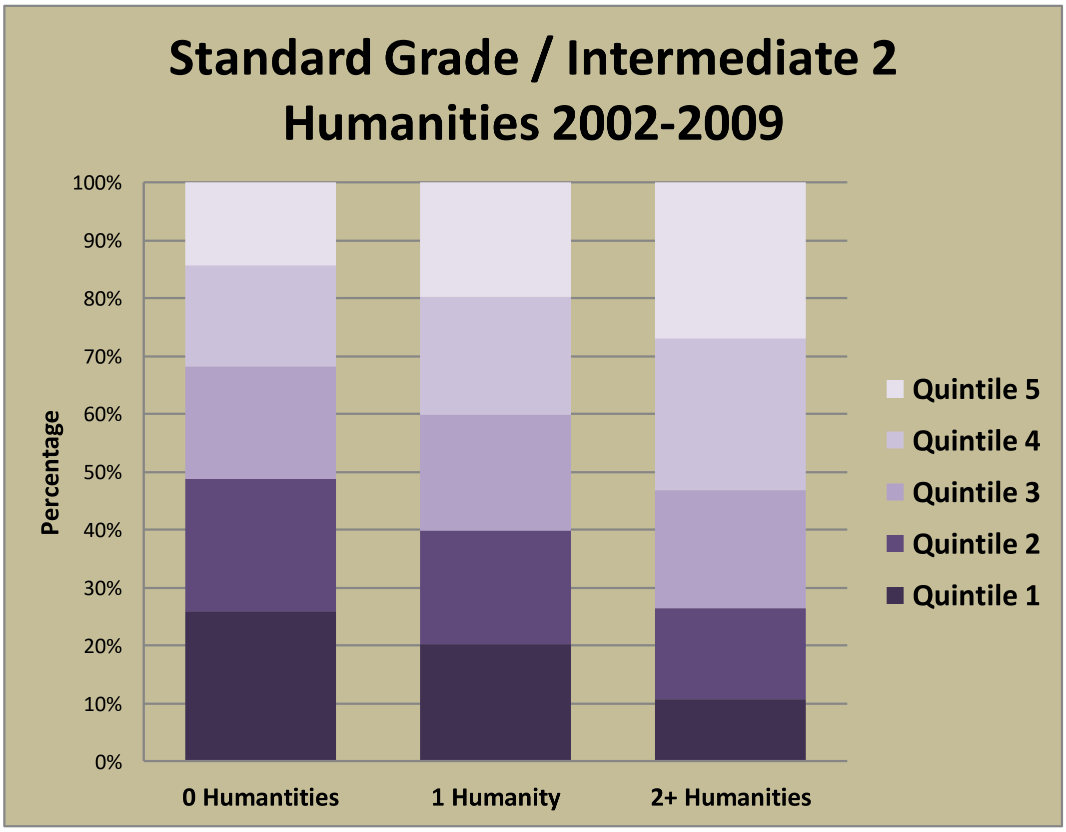
Pearson χ^2 (8) = 4800   Pr = 0.000

Figure A16: Number of Language Subjects Studied in S4 by SIMD



Pearson χ^2 (12) = 6500   Pr = 0.000

Figure A17: Number of Humanities Subjects Studied in S4 by SIMD



Pearson χ^2 (12) = 6800   Pr = 0.000

Appendix B

Appendix to Chapter Two

Table B1

Summary of Tests of Proportional Odds' Assumption for Facilitating Subjects'

Ordered Logit Models of Attainment

STEM			Humanities			Languages		
Qualification Level/Subject	**Test Statistic**	**Model**	**Qualification Level/Subject**	**Test Statistic**	**Model**	**Qualification Level/Subject**	**Test Statistic**	**Model**
Standard Grade / Intermediate 2								
Maths	2765.5***	MLM	English	2052.30***	MLM	French	2478.30***	MLM
Biology	1088.3***	MLM	Geography	1274.74***	MLM	German	927.89***	MLM
Chemistry	1157.81***	MLM	History	1376.89***	MLM	Spanish	4.14.64***	MLM
Physics	941.54***	MLM	Modern Studies	581.81***	MLM			
Computing	968.47***	MLM						
Higher								
Maths	468.42***	MLM	English	534.21***	MLM	French	244.28***	MLM
Biology	396.96***	MLM	Geography	212.36***	MLM	German	182.82***	MLM
Chemistry	250.72***	MLM	History	367.94***	MLM	Spanish	160.54***	MLM
Physics	223.16***	MLM	Modern Studies	206.89***	MLM			
Computing	255.01***	MLM						
Advanced Higher								
Maths	170.43***	MLM	English	211.21***	MLM	French	125.24	Ologit
Biology	206.92***	MLM	Geography	DNC		German	125.51*	MLM
Chemistry	136.63*	MLM	History	165.20***	MLM	Spanish	129.59***	MLM
Physics	120.02	Ologit	Modern Studies	156.71***	MLM			
Computing	175.05***	MLM						

*** p<0.001 ** p<0.01 * p<0.05

DNC = did not converge

Table B2

Sample Summary Statistics: Attainment Models

Total Observations		**Total No Cases**	**Observed once**	**Observed twice**	**Observed 3/+**
1,021,470		489,468	166,136	114,662	208,670
Dependent Variables					
SGI Grade Models	**Obs**	**Higher Grade Models**	**Obs**	**Adv Higher Grade Models**	**Obs**
English	378,254	English	109,015	English	6988
Geography	123,097	Geography	28,991	Geography	3950
History	135,516	History	33,967	History	4086
Modern Studies	76,520	Modern Studies	26,636	Modern Studies	3168
French	189,053	French	19,571	French	2288
German	63,272	German	7,779	German	755
Spanish	14,980	Spanish	3,013	Spanish	460
Maths	284,325	Maths	80,817	Maths	11016
Biology	133,430	Biology	45,556	Biology	7325
Chemistry	128,679	Chemistry	42,520	Chemistry	7205
Physics	104,383	Physics	39,480	Physics	5366
Independent Variables of Main Interest		**N**		**%**	
Sex (Male = 1, Female = 0)		M: 247,210	F: 242,258	M: 50.5%	F: 49.5%
SIMD of Household					
SIMD 1		102,211		21%	
SIMD 2		97,291		20%	
SIMD 3		95,861		20%	
SIMD 4		95,457		20%	
SIMD 5		90,073		19%	
Total SIMD Nos		480,893		2% Missing (8,575)	
Continuous Variables	**Mean**	**Standard Deviation**	**Maximum**	**Minimum**	**Obs**
SGI points	159	83	422	0	444,668
SGI Cohort Size	181	71	536	1	489,452
School % SIMD 1 & 2 pupils	39	25	100	0	489,468
Other Variables		**N**		**%**	
Urban/Rural Location of School attended					
Accessible Rural		55,186		11%	
Accessible Small Towns		43,556		9%	
Large Urban Areas		168,506		35%	
Other Urban Areas		161,031		34%	
Remote Rural		32,371		7%	
Remote Small Towns		20,243		4%	
Total		480,893		2% Missing (8,575)	
Local Authority of School attended					
Aberdeen City		15,743		3.5%	
Aberdeenshire		24,859		5.5%	
Angus		10,053		2%	

Table B2 cont.

Sample Summary Statistics: Attainment Models

Local Authority of School attended		
Argyll & Bute	8,497	2%
Clackmannanshire	2,941	0.7%
Dumfries & Galloway	15,743	3.5%
Dundee City	10,779	2%
East Ayrshire	12,150	3%
East Dunbartonshire	11,102	2.5%
East Lothian	8,802	2%
East Renfrewshire	11,084	2.5%
Edinburgh, City of	29,557	6.5%
Eilean Siar	2,906	0.6%
Falkirk	14,649	3%
Fife	35,440	8%
Glasgow, City of	42,929	9.5%
Highland	19,764	4%
Inverclyde	2,108	0.5%
Midlothian	9,453	2%
Moray	9,315	2%
North Ayrshire	12,371	3%
North Lanarkshire	31,575	7%
Orkney Islands	2,235	0.5%
Perth & Kinross	12,122	3%
Renfrewshire	16,068	3.5%
Scottish Borders	9,431	2%
Shetland Islands	2,572	0.6%
South Ayrshire	9,227	2%
South Lanarkshire	25,521	6%
Stirling	7,787	2%
West Dunbartonshire	9,588	2%
West Lothian	16,258	3.5%
Total	452,629	7% Missing (36,776)
Year Dummy		
2002	59,083	12%
2003	60,211	12%
2004	60,465	12%
2005	60,817	12%
2006	64,811	13%
2007	64,066	13%
2008	62,124	13%
2009	57,891	12%

Appendix C

Appendix to Chapter Three

Table C1

Sample Summary Statistics: STEM Subject Choice – Population Models

Total Observations		**Total No Cases**	**Observed once**	**Observed twice**	**Observed 3/+**
1,021,470		489,468	166,136	114,662	208,670
Dependent Variables					
SGI Choice Models	**Obs**	**Higher Choice Models**	**Obs**	**Adv Higher Choice Models**	**Obs**
Maths	444,900	Maths	185,698	Maths	53,676
Biology	444,900	Biology	185,698	Biology	53,676
Chemistry	444,900	Chemistry	185,698	Chemistry	53,676
Physics	444,900	Physics	185,698	Physics	53,676
Science	442706	Computing	185,698	Computing	53,200
Computing	444,900				
Independent Variables of Main Interest		**N**		**%**	
Sex (Male = 1, Female = 0)		M: 247,210	F: 242,258	M: 50.5%	F: 49.5%
SIMD of Household					
SIMD 1		102,211		21%	
SIMD 2		97,291		20%	
SIMD 3		95,861		20%	
SIMD 4		95,457		20%	
SIMD 5		90,073		19%	
Total SIMD Nos		480,893		2% Missing (8,575)	
Continuous Variables	**Mean**	**Standard Deviation**	**Maximum**	**Minimum**	**Obs**
SGI points	159	83	422	0	444,668
SGI Cohort Size	181	71	536	1	489,452
School % SIMD 1 & 2 pupils	39	25	100	0	489,468
Other Variables		**N**		**%**	
Urban/Rural Location of School attended					
Accessible Rural		55,186		11%	
Accessible Small Towns		43,556		9%	
Large Urban Areas		168,506		35%	
Other Urban Areas		161,031		34%	
Remote Rural		32,371		7%	
Remote Small Towns		20,243		4%	
Total		480,893		2% Missing (8,575)	
Local Authority of School attended					
Aberdeen City		15,743		3.5%	
Aberdeenshire		24,859		5.5%	
Angus		10,053		2%	
Argyll & Bute		8,497		2%	
Clackmannanshire		2,941		0.7%	
Dumfries & Galloway		15,743		3.5%	
Dundee City		10,779		2%	

Table C1 cont.

Sample Summary Statistics: STEM Subject Choice – Population Models

Local Authority of School attended		
East Ayrshire	12,150	3%
East Dunbartonshire	11,102	2.5%
East Lothian	8,802	2%
East Renfrewshire	11,084	2.5%
Edinburgh, City of	29,557	6.5%
Eilean Siar	2,906	0.6%
Falkirk	14,649	3%
Fife	35,440	8%
Glasgow, City of	42,929	9.5%
Highland	19,764	4%
Inverclyde	2,108	0.5%
Midlothian	9,453	2%
Moray	9,315	2%
North Ayrshire	12,371	3%
North Lanarkshire	31,575	7%
Orkney Islands	2,235	0.5%
Perth & Kinross	12,122	3%
Renfrewshire	16,068	3.5%
Scottish Borders	9,431	2%
Shetland Islands	2,572	0.6%
South Ayrshire	9,227	2%
South Lanarkshire	25,521	6%
Stirling	7,787	2%
West Dunbartonshire	9,588	2%
West Lothian	16,258	3.5%
Total	452,629	7% Missing (36,776)
Year Dummy		
2002	59,083	12%
2003	60,211	12%
2004	60,465	12%
2005	60,817	12%
2006	64,811	13%
2007	64,066	13%
2008	62,124	13%
2009	57,891	12%

Table C2

Sample Summary Statistics: STEM Subject Choice & Attainment – Twins/CSS Models

Total Observations	Total No Cases	Observed once	Observed twice	Observed 3/+
24,879	11,921	4,006	2,872	5,043

Dependent Variables

SGI Choice Models	Obs	Higher Choice Models	Obs	Adv Higher Choice Models	Obs
Maths	10872	Maths	4347	Maths	1251
Biology	10872	Biology	4326	Biology	1250
Chemistry	10872	Chemistry	4347	Chemistry	1182
Physics	10872	Physics	4347	Physics	1213
Science	10832	Computing	4324	Computing	917
Computing	10872				
SGI Attainment Models	**Obs**	**Higher Attainment Models**	**Obs**	**Adv Higher Attainment Models**	**Obs**
Maths	8767	Maths	1926	Maths	DNC [+]
Biology	3444	Biology	1110	Biology	DNC
Chemistry	3226	Chemistry	1024	Chemistry	DNC
Physics	2507	Physics	892	Physics	DNC
Science	1065	Computing	450	Computing	DNC
Computing	2625				

Independent Variables of Main Interest	N		%	
Sex (Male = 1, Female =0)	M: 5,876	F: 6,045	M: 50.5%	F: 49.5%
SIMD of Household				
SIMD 1	2,420		21%	
SIMD 2	2,671		23%	
SIMD 3	2,365		20%	
SIMD 4	2,140		18%	
SIMD 5	2,198		19%	
Total SIMD Nos	11,794		0% Missing	

Continuous Variables	Mean	Standard Deviation	Maximum	Minimum	Obs
SGI points	160	84	378	0	10,771
SGI Cohort Size	182	73	536	1	11,921
School % SIMD 1 & 2 pupils	38	25	100	0	11,921

Other Variables	N	%
Urban/Rural Location of School attended		
Accessible Rural	1,267	10.7%
Accessible Small Towns	994	8.4%
Large Urban Areas	4,479	38.0%
Other Urban Areas	3,911	33.2%
Remote Rural	718	6.1%
Remote Small Towns	425	3.6%
Total	11,794	0% Missing

Table C2 cont.

Sample Summary Statistics: STEM Subject Choice & Attainment – Twins/CSS Models

Local Authority of School attended		
Aberdeen City	446	4.1%
Aberdeenshire	552	5.0%
Angus	264	2.4%
Argyll & Bute	237	2.2%
Clackmannanshire	74	0.7%
Dumfries & Galloway	351	3.2%
Dundee City	255	2.3%
East Ayrshire	289	2.6%
East Dunbartonshire	281	2.6%
East Lothian	215	2.0%
East Renfrewshire	309	2.8%
Edinburgh, City of	745	6.8%
Eilean Siar	74	0.7%
Falkirk	359	3.3%
Fife	826	7.5%
Glasgow, City of	993	9.0%
Highland	446	4.1%
Inverclyde	61	0.6%
Midlothian	202	1.8%
Moray	190	1.7%
North Ayrshire	261	2.4%
North Lanarkshire	826	7.5%
Orkney Islands	40	0.4%
Perth & Kinross	302	2.8%
Renfrewshire	396	3.6%
Scottish Borders	231	2.1%
Shetland Islands	60	0.6%
South Ayrshire	208	1.9%
South Lanarkshire	676	6.2%
Stirling	199	1.8%
West Dunbartonshire	246	2.2%
West Lothian	367	3.3%
Total	10,981	8% Missing (940)
Year Dummy		
2002	1,265	10.6%
2003	1,318	11.1%
2004	1,376	11.5%
2005	1,412	11.8%
2006	1,677	14.1%
2007	1,649	13.8%
2008	1,776	14.9%
2009	1,448	12.2%

[+] Notes: DNC = Did not converge. No AH female only models converged, only Maths and Biology converged for Twins/CSS whole sample.

Table C3

Summary of Classification/Goodness of Fit Tests for Logistic Regression Models of Subject Choice

	Full Dataset		**Twins/CSS All**		**Twins/CSS Female**	
Qualification Level/Subject	**Correctly Classified / Error Rate Reduction**	**GoF**	**Correctly Classified / Error Rate Reduction**	**GoF**	**Correctly Classified / Error Rate Reduction**	**GoF**
Standard Grade / Intermediate 2						
Maths	82.16% -7%	127515.97***	81.95% -7%	9701.26***	82.01% -7%	4849.69***
Biology	70.63% -8%	53436.57***	71.80% -11%	8401.93***	63.53% -17%	4407.01***
Chemistry	70.20% -1%	51483.84***	70.47% -0.5%	8491.64***	70.35% -2%	4336.16***
Physics	76.14% -2%	46354.40***	77.70% -3%	8458.88***	86.96% 0%	4250.06***
Science	90.27% -1%	81536.23***	90.11% -0.5%	9039.57***	90.44% -1%	4705.79***
Computing	74.36% -1%	66751.54***	76.23% -2%	8573.99***	82.30% 0%	4405.46***
Higher						
Maths	79.34% -52%	287813.60***	80.08% -55%	7975.05***	80.10% -52%	2476.37***
Biology	76.15% -117%	159236.76***	76.84% -10%	4469.75***	70.56% -14%	2196.18
Chemistry	79.16% -10%	176009.10***	80.10% -16%	4355.44***	80.04% -14%	2356.05*
Physics	82.27% -17%	175278.33***	83.28% -19%	4886.30***	88.97% -2%	2264.84
Computing	89.57% 0%	158172.11***	89.64% 0%	3819.51	93.56% -1%	1674.05

Table C3 cont.

Summary of Classification/Goodness of Fit Tests for Logistic Regression Models of Subject Choice

	Full Dataset		**Twins/CSS All**		**Twins/CSS Female**	
Qualification Level/Subject	**Correctly Classified / Error Rate Reduction**	**GoF**	**Correctly Classified / Error Rate Reduction**	**GoF**	**Correctly Classified / Error Rate Reduction**	**GoF**
Advanced Higher						
Maths	79.49% 0%	44401.35				
Biology	86.35% 0%	45019.82*				
Chemistry	86.58% 0%	43996.98				
Physics	89.99% 0%	42899.32				
Computing	95.93% 0%	42976.03				

*** p<0.001 ** p<0.01 * p<0.05

DNC = did not converge

Table C4

Summary of Tests of Proportional Odds' Assumption for Twins/CSS' Ordered Logit Models of Attainment

	Twins/CSS All		**Twins/CSS Female**	
Qualification Level/Subject	**Test Statistic**	**Model**	**Test Statistic**	**Model**
Standard Grade / Intermediate 2				
Maths	217.77***	MLM	183.51***	MLM
Biology	177.06***	MLM	190.24***	MLM
Chemistry	165.44***	MLM	146.39***	MLM
Physics	123.49	Ologit	140.99**	MLM
Science	431.20***	MLM	290.17***	MLM
Computing	151.94**	MLM	162.47***	MLM
Higher				
Maths	145.79*	MLM	DNC	
Biology	130.45	Ologit	113.04	Ologit
Chemistry	141.18*	MLM	162.52***	MLM
Physics	129.94	Ologit	134.51**	MLM
Computing	DNC		173.28***	MLM
Advanced Higher				
Maths	DNC		DNC	
Biology	DNC		146.15***	MLM
Chemistry	DNC		122.55***	MLM
Physics	144.76***	MLM	DNC	
Computing	DNC		DNC (N=2)	

*** $p<0.001$ ** $p<0.01$ * $p<0.05$

DNC = did not converge

Bibliography

Anders, J., Henderson, M., Moulton, V. & Sullivan, A. (2018), The role of schools in explaining individuals' subject choices at age 14, *Oxford Review of Education*, Vol. 44, No. 1, pp75-93.

Andrews, J., Robinson, D. & Hutchinson, J. (2017), *Closing the Gap? Trends in Educational Attainment and Disadvantage*, Education Policy Institute, August.

Angrist, J. D. & Pischke, J. (2008), *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press.

Archer, L., Osborne, J. F., Dillon, J. S., DeWitt, J., Willis, B., & Wong, B. (2013), *Interim Research Summary, ASPIRES Project: What shapes children's science and career aspirations age 10-13?*, King's College London.

Auyeung, B., Lombardo, M. V. & Baron-Cohen, S. (2013), Prenatal and postnatal hormone effects on the human brain and cognition, *Pflügers Archiv-European Journal of Physiology*, Vol. 465, No. 5, pp557–571.

Ayorech, Z., Krapohl, E., Plomin, R. & von Stum, S. (2017), Genetic Influence on Intergenerational Educational Attainment, *Psychological Science*, Vol. 28, No. 9, pp1302 – 1310.

Becker, G. S. (1993) [1964], *Human capital: a theoretical and empirical analysis, with special reference to education,* 3rd ed, Chicago: The University of Chicago Press.

Bernardi, F. (2012), Unequal transitions: Selection bias and the compensatory effect of social background in educational careers, *Research in Social Stratification and Mobility*, Vol. 30, pp159-174.

Bhalotra, S. & Clarke, D. (2018), The Twin Instrument: Fertility and Human Capital Investment, IZA Discussion Paper No. 11878, October

Blanden, J., Gregg, P. & Machin, S. (2003), *Changes in Educational Inequality*, The Centre for Market and Public Organisation (CMPO) Working Paper Series No. 03/079.

Blanden, J., Goodman, A., Gregg, P. & Machin, S. (2004), 'Changes in Intergenerational Mobility in Britain', in Corak, M. (Editor), *Generational Income Mobility in North America and Europe*, Cambridge University Press.

Blanden, J., & Gregg, P. (2004), Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain, *Oxford Review of Economic Policy*, Vol. 20, No. 2, pp.245-263.

Blundell, R., Dearden, L., and Sianesi, B. (2005), Evaluating the impact of education on earnings: Models, methods and results from the NCDS, *Journal of the Royal Statistical Society Series A*, Vol. 168, No. 3, pp473-512.

Boaler, J., William, D. & Brown, M. (2000), Students' experiences of ability grouping – disaffection, polarisation and the construction of failure, *British Educational Research Journal*, Vol. 26, No. 5, pp631-64.

Bogard, M. (2013), *Selection Bias and the Rubin Causal Model and Potential Outcomes Framework*, http://econometricsense.blogspot.com/2013/05/selection-bias-and-rubin-causal-model.html, accessed 08/02/2019.

Bonjour, D., Cherkas, L, Haskel, J., Hawkes, D. & Spector, T. (2002), *Returns to Education: Evidence from UK Twins*, Centre for Economics for Education, April.

Boudon, R. (1974), *Educational Opportunity and Social Inequality: Changing Prospects in Western Society*, New York, John Wiley & Sons.

Bradshaw, P. (2011), *Growing up in Scotland: Changes in child cognitive ability in the pre-school years*, Scottish Government, Edinburgh.

Breen, R. & Goldthorpe, J. H. (1997), Explaining Educational Differentials: Towards a Formal Rational Action Theory, *Rationality and Society*, Vol. 9, No. 3, pp275-305.

Breen, R. & Jonsson, J. O. (2000), Analyzing Educational Careers: A Multinomial Transition Model, *American Sociological Review*, Vol.65, No. 5, pp754-772.

Breen, R., Luijkx, R., Muller, W., Pollak, R. (2009), Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries, *American Journal of Sociology*, Vol. 114, No. 5 (March), pp. 1475-1521.

Broughton, N. (2013), *In the Balance: the STEM Human Capital Crunch*, The Social Market Foundation, March.

Burgess, S., McConnell, B., Propper, C. & Wilson, D. (2003), *Girls Rock, Boys Roll: An Analysis of the Age 14-16 Gender Gap in English Schools*, Centre for Market and Public Organisation, (CMPO) Working Paper Series No. 03/084.

Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H., et al (2012), Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children, *Behavior Genetics*, Vol. 42, pp699–710 [PubMed: 22700061].

Cameron, S. V. & Heckman, J. J., (1998), Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Male, *Journal of Political Economy*, Vol. 106, No. 2, pp262-333, April.

Cameron, A. C. & Miller, D. L. (2015), A Practioner's Guide to Cluster-Robust Inference, *Journal of Human Resources*, Vol. 50, No. 2, pp317-372, March.

Card, D. (1999), 'The Causal Effect of Education on Earnings', in Ashenfelter, O. & Card, D., (eds.), *Handbook of Labor Economics*, Elsevier Science B.V..

Chevalier, A. & Lanot, G. (2002), The Relative Effect of Family Characteristics and Financial Situation on Educational Achievement, *Education Economics*, Vol. 10, No. 2.

Chou, S., Liu, E., Lin, M., & Liu, J. (2015), Better peers, better scores? A study of twin junior high school graduates in Taiwan, *Applied Economics*, Vol. 47, No. 32, pp3462-3481.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. & Vignoles, A. (2013), Widening participation in higher education: analysis using linked administrative data, *Journal of the Royal Statistical Society*, Vol. 176, No. 2, pp431-457.

Christensen, K., Petersen, I., Skytthe, A., Herskind, A. M., McGue, M., Bingley, P. (2006), Comparison of academic performance of twins and singletons in adolescence, *British Medical Journal*, 10.1136/bmj.38959.650903.

Cohen, C. C. C., van Goozen, S. H. M., Orlebeke, J. F., Buitelaar, J. K. & Cohen-Kettenis, P.T. (2002), A Comparison of Educational Achievement in a National Sample of Dutch Female Twins and Their Matched Singleton Controls, *Twin Research*, Vol. 5 No. 4, pp273–277, August.

Cohen-Bendahan, C. C. C., Buitelaar, J. K., van Goozen, S. H. M., & Cohen-Kettenis, P. T. (2004), Prenatal exposure to testosterone and functional cerebral lateralization: a study in same-sex and opposite-sex twin girls, *Psychoneuroendocrinology*, Vol. 29, No. 7, pp911–916 in Gielen, A. C. & Zwiers, E. S. (2018), *Biology and the gender gap in educational performance - The role of prenatal testosterone in test scores*, IZA Discussion Paper No. 11936, November.

Commission on School Reform (2013), *By Diverse Means: Improving Scottish Education*, Centre for Scottish Public Policy & Reform Scotland, March.

Commission on Widening Access (2015), *Interim Report*, The Scottish Government, November.

Commission on Widening Access (2016), *A Blueprint for Fairness: The Final Report of the Commission on Widening Access*, The Scottish Government, March.

Connolly, P. (2006), The effects of social class and ethnicity on gender differences in GCSE attainment: a secondary analysis of the Youth Cohort Study of England and Wales 1997–2001, *British Educational Research Journal*, Vol. 32, No. 1, February, pp3-21.

Cronqvist, H., Previtero, A., Siegel, S. & White, R. E. (2016), The fetal origins hypothesis in finance: Prenatal environment, the gender gap, and investor behaviour, *The Review of Financial Studies*, Vol. 29, No. 3, pp739–786, March.

Croxford, L. (2006), *The Youth Cohort Surveys – How good is the evidence?*, Special CES Briefing No. 38, Centre for Educational Sociology, May.

Croxford, L. (2009), *Change Over Time in the Context, Outcomes and Inequalities of Secondary Schooling in Scotland, 1985-2005*, Education Analytical Services, Scottish Government.

Croxford, L., Ozga, J. and Provan, F. (2004), *Education Maintenance Allowances (EMAs): Evaluation of the East Ayrshire Pilot, Stage 2*, University of Edinburgh: CES.

Croxford, L. & Raffe, D., (2007), 'Education Markets and social class inequality: a comparison of trends in England, Scotland and Wales', in Teese, R., Lamb, S. & Duru-Bellat, M. (eds), *Education and Inequality: International Perspectives on Theory and Policy, Volume 3,* Springer.

Dearden, L. (2010), *Administrative Data and Economic Policy Evaluation*, Department of Quantitative Social Science Working Paper No. 10-14, Institute of Education, University of London, July.

Dearden, L., McIntosh, S., Mychk, M. & Vignoles, A. (2002), The returns to academic and vocational qualifications in the UK, *Bulletin of Economic Research*, Vol 54, No. 3, pp249-274.

Dearden, L., McGranahan, L., & Sianesi, B. (2004), *An in-depth analysis of the returns to National Vocational Qualifications obtained at Level 2*, Centre for the Economics of Education (CEE) Discussion Paper No.46.

Dearden, L., Micklewright, J. & Vignoles, A. (2011), The Effectiveness of English Secondary Schools for Pupils of Different Ability Levels, *Fiscal Studies*, Vol. 32, No. 2, June.

Deary, I. J., Pattie A., Wilson, V., Whalley, L. J. (2005), The cognitive cost of being a twin: two whole-population studies, *Twin Res Hum Genet*; Vol. 8, pp376-383.

Deary, I. J. (2006), Educational performance in twins: Is no different from that seen in singletons by adolesence, *British Medical Journal*, Vol. 333, pp1080-1081, November.

Deary I. J., Strand, S., Smith, P., Fernandes, C. (2007), Intelligence and educational achievement, *Intelligence*, Vol. 35, No.1, pp13-21.

De Philippis, M. (2016), STEM Graduates and Secondary School Curriculum: Does Early Exposure to Science Matter?, Centre for Economic Performance Discussion Paper No. 1443, August, CEP.

de Zeeuwa, E. L., van Beijsterveldta, C. E.M., Glasnera, T. J., de Geusa, E. J. C. & Boomsa, D. I. (2016), Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children, *Learning and Individual Differences – Journal of Psychology and Education*, Elsevier, Vol. 47, pp156–166.

de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

Erikson, R., Goldthorpe, J. H., Jackson, M., Yaish, M. and Cox, D. R. (2005), On Class Differentials in Educational Attainment, *Proceedings of the National Academy of Sciences*, Vol. 102, pp9730-9733.

Erikson, R. & Jonsson, J.O. (1996), The Swedish context: educational reform and long-term change in educational inequality, in Erikson, R. & Jonsson, J.O. (eds.) (1996), *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, pp65–94. Westview, Boulder.

Evans, D.M. & Martin, N. G. (2000), The validity of twin studies, *GeneScreen*, Wiley, Vol 1, No. 2, pp77–79.

Forde, C., Kane, J., Condie, R., McPhee, A. & Head, G. (2006), *Strategies to Address Gender Inequalities in Scottish Schools A Review of the Literature*, Information and Analytical Services Division, Scottish Executive Education Department, May.

Frey, M. C. & Detterman, D. K., (2004), Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability, *Psychological Science*, Vol. 15, pp373–378.

Gasteen, A., Houston, J. (2007), Employability and Earnings Returns to Qualifications in Scotland, *Regional Studies*, Vol. 41, No. 4, pp443-452.

Gayle, V., Berridge, D. & Davies, R. (2002), Young People's Entry into Higher Education: quantifying influential factors, *Oxford Review of Education*, Vol. 28, No. 1.

Gayle, V. & Playford, C. J. (2014), *The concealed middle? An exploration of ordinary young people and school GCSE subject area attainment*, ESRC Centre for Population Change Working Paper 51, June.

Gayle, V., Playford, C. J., Connelly, R. & Murray, S. (2016), *Between the NEET and the tidy – Exploring 'middle' outcomes in Scottish school qualifications*, ESRC Centre for Population Change Working Paper 76, March.

Gielen, A. C. & Zwiers, E. S. (2018), *Biology and the gender gap in educational performance - The role of prenatal testosterone in test scores*, IZA Discussion Paper No. 11936, November.

Gielen, A. C., Holmes, J. & Myers, C. (2016), Prenatal testosterone and the earnings of men and women, *Journal of Human Resources*, Vol. 51, No. 1, pp30–61.

Goodman, A. & Gregg, P. (2010), *Poor children's educational attainment: how important are attitudes and behaviour?*, Joseph Rowntree Foundation, March.

Gorard, S., Rees, G. & Salisbury, J. (1999), Reappraising the apparent underachievement of boys at school, *Gender and Education*, Vol. 11, No. 4, pp441-445.

Gregg, P. and Tominey, E. (2005). The Wage Scar from Male Youth Unemployment, *Labour Economics*, Vol.12, No. 4, pp487-509.

Gujarati, D. (2015), *Econometrics by Example*, Second Edition, Palgrave.

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S. & Plomin, R. (2012), Socioeconomic Status (SES) and Children's Intelligence (IQ): In a UK-Representative Sample SES Moderates the Environmental, Not Genetic, Effect on IQ, PLoS ONE, Vol. 7, No. 2: e30320. https://doi.org/10.1371/journal.pone.0030320.

Harlaar, N, Spinath, F. M., Dale, P. S. & Plomin, R. (2005), Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins, *Journal of Child Psychology and Psychiatry*, Vol. 46, pp373–384 in de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

Hastings, S. (2006), *White underachievement*, TES, 28 April, https://www.tes.com/news/white-underachievement accessed 26/07/2018.

Hauser, R. M. & Andrew, M. (2006), Another Look at the Stratification of Educational Transitions: The Logistic Response Model with Partial Proportionality Constraints, *Sociological Methodology*, Vol. 36, pp1-26.

Haworth, C. M. A., Dale, P. S. & Plomin, R. (2008), A Twin Study into the Genetic and Environmental Influences on Academic Performance in Science in nine-year-old Boys and Girls, *International Journal of Science Education*, Vol. 30, No. 8, pp1003-1025.

Haworth, C. M. A., Dale, P. S., & Plomin, R. (2009), The Etiology of Science Performance: Decreasing Heritability and Increasing Importance of the Shared Environment From 9 to 12 Years of Age, *Child Development*, Vol. 80, No. 3, pp662–673, May/June.

Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C. E. et al (2010), The heritability of general cognitive ability increases linearly from childhood to young adulthood, *Molecular Psychiatry*, Vol. 15, No. 11, pp1112–1120 in Vinkhuyzen, A. A., van der Sluis, S., Maes, H. H, Posthuma, D. (2012), Reconsidering the heritability of intelligence in adulthood: Taking assortative mating and cultural transmission into account, *Behavior Genetics*, Vol. 42, No. 2, pp187–198.

Haworth, C. M. A., Asbury, K., Dale, P. S. & Plomin, R. (2011), Added Value Measures in Education Show Genetic as Well as Environmental Influence, PLoS ONE Vol. 6, No.2: e16006, February, doi:10.1371/journal.pone.0016006.

Heath, A. C., Berg, K., Eaves, L. J., Solaas, M. H., Corey, L. A., Sundet, J., Magnus, P. & Nance, W. E., (1985), Education policy and the heritability of educational attainment, *Nature*, Vol. 314(6013), pp734–736 in de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

Heil, M., Kavšek, M., Rolke, B., Beste, C. &, Jansen, P. (2011), Mental rotation in female fraternal twins: Evidence for intra-uterine hormone transfer?, *Biol Psychol,* Vol 86, No. 1, pp90-93, January.

Hillman, N. & Robinson, N. (2016), *Boys to Men: The underachievement of young men in higher education – and how to start tackling it*, Higher Education Polich Institute (HEPI) Report 84, May.

Hills, J., Brewer, M., Jenkins, S., Lister, R., Lupton, R., Machin, S., Mills, C., Modood, T., Rees, T. & Riddell, S. (2010), *An Anatomy of Economic Inequality in the UK, report of the National Equality Panel*, London: Government Equalities Office.

Holm, A. & Jæger, M. M. (2011), Dealing with selection bias in educational transition models: The bivariate probit selection model, *Research in Social Stratification and Mobility*, Vol. 29, pp311-322.

Howieson, C. & Iannelli, C. (2008), The effects of low attainment on young people's outcomes at age 22-23 in Scotland, *British Educational Research Journal*, Volume 34, Issue 2, pages 269–290, April.

Hu, A. (2017), Using a discretized measure of academic performance to approximate primary and secondary effects in inequality of educational opportunity, *Quality & Quantity*, Vol. 51, No. 4 pp1627-1643, (July).

Hupkau, C., McNally, S., Ruiz-Valenzuela, J. & Ventura, G. (2016), *Post-Compulsory Education in England: Choices and Implications*, Centre for Vocational Education Research, Discussion Paper 001, July.

Iannelli, C. (2013), The role of school curriculum in social mobility, *British Journal of Sociology of Education*, Vol. 34, No. 5/6, pp907–928.

Iannelli, C. & Klein, M. (2015), *Subject choice and inequalities in access to Higher Education: Comparing Scotland and Ireland*, AQMeN Research Briefing 7, March.

Iannelli, C., Smythe, E. & Klein, M. (2016), Curriculum differentiation and social inequality in higher education entry in Scotland and Ireland, *British Educational Research Journal*, Vol. 42, No. 4, August, pp. 561–581.

Imbens, G. W. & Wooldridge, J. M.(2009), Recent developments in the econometrics of program evaluation, *Journal of Economic Literature*, Vol. 47, No. 1, pp. 5–86.

Institute of Physics (2018), *Why not Physics? A snapshot of girls' uptake at A-level*, Institute of Physics, May.

Jackson, M. (2013a) Introduction: how is inequality of educational opportunity generated? the case for primary and secondary effects, in Jackson, M. (ed) *Determined to Succeed? Performance Versus Choice in Educational Attainment*, Stanford University Press, Stanford, pp1–33.

Jackson, M. (2013b) *Determined to Succeed? Performance Versus Choice in Educational* Attainment, Stanford University Press, Stanford.

Jackson, M., Erikson, R., Goldthorpe, J. H. & Yaish, M. (2007), Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to A-Level Courses in England and Wales, *Acta Sociologica*, Vol. 50, No.3, pp211–229, September.

Johnson, W., McGue, M., & Iacono, W. G. (2006), Genetic and environmental influences on academic achievement trajectories during adolescence, *Dev Psychology,* Vol. 42, No.3, pp514–532.

Jones, J., Joyce, M. & Thomas, J (2003), Non-employment and labour availability, *Bank of England Quarterly Bulletin*, Bank of England, London.

Jordan-Young, R. M. (2010). Brain storm. Harvard University Press in Gielen, A. C. & Zwiers, E. S. (2018), *Biology and the gender gap in educational performance - The role of prenatal testosterone in test scores*, IZA Discussion Paper No. 11936, November.

Karlson, K. B. (2011), Multiple paths in educational transitions: A multinomial transition model with unobserved heterogeneity, *Research in Social Stratification and Mobility*, Vol. 29, pp323-341

Kartsonaki, C., Jackson, M., Cox, D.R. (2013), Primary and secondary effects: some methodological issues, in Jackson, M. (ed) Determined to Succeed? Performance Versus Choice in Educational Attainment, Stanford University Press, Stanford pp. 34–55.

Kim, A. & Kim, K. (2003), *Returns to Tertiary Education in Germany and the UK: Effects of Fields of Study and Gender*, Working Paper No. 62, Mannheim Centre for European Social Research, University of Mannheim.

Klein, M. & Iannelli, C. (2014), *Degrees of difference: Social inequalities in graduates' job opportunities in the UK and Germany*, AQMeN Research Briefing 6, November.

Kohler, U. & Krueter, F. (2009), *Data Analysis Using STATA*, Second edition, STATA Press, Texas.

Kovas, Y., Haworth, C. M., Petrill, S. A. & Plomin, R. (2007), Mathematical ability of 10-year-old boys and girls: genetic and environmental etiology of typical and low performance, *Journal of Learning* Disabilities, Vol. 40, pp554–567 in de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S.B., Dale, P.S. & Plomin, R. (2013), Literacy and numeracy are more heritable than intelligence in primary school, *Psychological Science*, Vol. 24, pp2048–2056.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J., Asbury, K., Harlaar, N., Kovas,Y., Dale, P. S. & Plomin, R. (2014), The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence, *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 111, No. 42, pp15273–15278.

Leckie, G., Pillinger, R., Jenkins, J. & Rasbash, J. (2010), School, family, neighbourhood: which is most important to a child's education?, *Significance*, The Royal Statistical Society, pp67-70, June.

Long, J. S. & Freese, J. (2006), *Regression Models for Categorical Dependent Variables Using STATA,* 2nd edition, STATA Press.

Lucas, S.R. (2001), Effectively maintained inequality: Education transitions, track mobility, and social background effects, *American Journal of Sociology*, Vol. 106, No.6, pp1642–1690.

Machin, S. & Pekkarinen, T. (2008), Global sex differences in test score variability, *Science*, Vol. 322, No. 5906, pp1331-1332, November.

Mare, R. D. (1979), Social background composition and educational growth, *Demography*, 16, 55-71.

Mare, R. D. (1980), Social background and school continuation decisions, *Journal of the American Statistical Association*, 75, 292-305.

Mare, R. D. (1981), Change and stability in educational stratification, *American Sociological Review*, Vol. 46, No. 1, pp72-87, Feb.

Mare, R. D. (2006), Response: Statistical Models of Educational Stratification – Hauser and Andrew's Models for School Transitions, *Sociological Methodology*, Vol. 36, pp27-37.

McIntosh, S. (2006), Further Analysis of the Returns to Academic and Vocational Qualifications, *Oxford Bulletin of Economics and Statistics*, Vol. 68, No. 2.

McIntosh, S., (2007), *A Cost–Benefit Analysis of Apprenticeships and Other Vocational Qualifications*, Research Report RR834, Department for Education and Skills.

Miller, P., Mulvey, C. & Martin, N. (2001), Genetic and environmental contributions to educational attainment in Australia, *Economics of Education Review*, Vol. 20, pp211-224.

Mood, C. (2010), Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It, *European Sociological Review*, Vol 26, No. 1, pp67–82.

Morelli, C.J. & Seaman, P. (2010), Devolution as a policy crucible: the case of universal free school meals, *Poverty & Public Policy*, Vol. 2, No. 1, pp.139-61.

Moulton, V., Sullivan, A., Henderson, M. & Anders, J. (2018), Does what you study at age 14-16 matter for educational transitions post-16?, Oxford Review of Education, Vol. 44, No. 1, pp94-117.

Murray, S. J. (2011), *Growing up in the 1990s: Tracks and Trajectories of the 'Rising 16s': A longitudinal analysis using the British Household Panel Survey*, School of Applied Social Sciences, University of Stirling, Stirling.

Nicoletti, C. & Rabe, B. (2010), *Inequality in Pupils' Educational Attainment: How Much Do Family, Sibling Type and Neighbourhood Matter?,* Institute for Social & Economic Research Working Paper No. 2010-26, July.

OECD (2007), *Review of the Quality and Equity of Education Outcomes in Scotland: Diagnostic Report*, OECD Publishing, Paris.

OECD (2011), An Overview of Growing Income Inequalities in OECD Countries: Main Findings, in *Divided We Stand: Why Inequality Keeps Rising*, OECD Publishing, Paris, pp21-44, https://doi.org/10.1787/9789264119536-3-en.

OECD (2015), *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, PISA, OECD Publishing, Paris.

Ørstavik, R.E., Czajkowski, N., Røysamb, E., Knudsen, G. P., Tambs, K. & Reichborn-Kjenneru, T. (2014), Sex Differences in Genetic and Environmental Influences on Educational Attainment and Income, *Twin Research and Human Genetics*, Vol. 17, No. 6, pp516–525.

Paterson, L. and Iannelli, C. (2007), Social Class and Educational Attainment: A Comparative Study of England, Wales and Scotland, *Sociology of Education*, Vol. 80, pp330–358, October.

Paterson, L. Pattie, A. & Deary, I. J. (2011), Social class, gender and secondary education in Scotland in the 1950s,*Oxford Review of Education*, Vol. 37, No. 3, pp383-401, August.

Peper, J. S., Brouwer, R. M., Van Baal, G. C. M., Schnack, H. G., Van Leeuwen, M., Boomsma, D. I., Kahn, R. S., & Pol, H. E. H. (2009), Does having a twin brother make for a bigger brain? *European Journal of Endocrinology*, Vol. 160, No. 5, pp739–746.

Perry, E. & Francis, B. (2010), *The Social Class Gap For Educational Achievement: a review of the literature*, RSA, December.

Petrill, S. A. & Thompson, L. A. (1994), The effect of gender upon heritability and common environmental estimates in measures of scholastic achievement, *Personal and Individual Differences*, Vol. 16, pp631–640 in de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

Pirrie, A. & Hockings, E. (2012), *Poverty, educational attainment and achievement in Scotland: a critical review of the literature*, Scotland's Commissioner for Children and Young People, Edinburgh.

Posthuma, D., De Geus, E. J. C., Bleichrodt, N. & Boomsma, D. I. (2000), Twin-singleton differences in intelligence? *Twin Res Hum Genet*; Vol. 3, pp83-87.

Purcell, S. (2002), Variance components models for gene-environment interaction in twin analysis, *Twin Research*, Vol. 5, pp554–571.

Raffe, D. (2000), Investigating the Education Systems of the United Kingdom, in Phillips, D. (Ed.) *The Education Systems of the United Kingdom*, Oxford, Symposium Books Ltd.

Raffe, D., Croxford, L., Iannelli, C., Shapira, M., Howieson, C. (2006), *Social-Class Inequalities in Education in England and Scotland*, Centre for Educational Sociology Special Briefing, No. 40, University of Edinburgh, May.

Raffe, D., Howieson, C. & Tinklin, T. (2005), The Introduction of a Unified System of Post-Compulsory Education in Scotland, *Scottish Educational Review*, 37(1), pp.46-57.

Raffe, D., Howieson, C. & Tinklin, T. (2007), The Impact of a Unified Curriculum and Qualifications System: The Higher Still Reform of Post-16 Education in Scotland, *British Educational Research Journal*, 33(4), pp.479-508.

Rasbash, J., Leckie, G., Pillinger, R. & Jenkins, J. (2010), Children's educational progress: partitioning family, school and area effects, *Journal of the Royal Statistical Society*, Series A, 173, pp657–682.

Reynolds, C. A., Hewitt, J. K., Erickson, M. T., Silberg, J. L., Rutter, M., Simonoff, E. et al. (1996), The genetics of children's oral reading performance, Journal of Child Psychology and Psychiatry, Vol. 37, 425–434 in de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

Rietveld, C. A. et al. (2013), GWAS of 126,559 individuals identifies genetic variants associated with educational attainment, *Science*, Vol. 340, No. 6139, pp1467-1471, June.

Ronalds, G. A., DeStavola, B. L., Leon, D.A. (2005), The cognitive cost of being a twin: evidence from comparisons within families in the Aberdeen children of the 1950s cohort study, *British Medical Journal*, Vol. 331, pp1306-1310.

Royal Society of Edinburgh (2012), *Tapping all our Talents: Women in science, technology, engineering and mathematics: a strategy for Scotland*, Royal Society of Edinburgh.

Rubin, D. B. (1974), Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology,* Vol. 66, No. 5, Oct., pp688-701.

Russell Group (2013), *Informed Choices*.

Scarr-Salapatek, S. (1971), Race, social class, and IQ, *Science*, Vol. 174, No.4016, pp1285–1295 in de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

Scott, J. (2015), *The Governance of Curriculum for Excellence in Scottish Secondary Schools: Structural Divergence, Curricular Distortion and Reduced Attainment*, written submission to the OECD Review of Curriculum for Excellence, https://www.academia.edu/20171586/OECD_Evidence_Paper_2015.

Scottish Education Department (1977), *Assessment for All: Report of the Committee to Review Assessment in the Third and Fourth Years of Secondary Education in Scotland*, Edinburgh, HMSO.

Scottish Government (2016), *Scottish Survey of Literacy and Numeracy 2015 (Numeracy)*, May, http://www.gov.scot/Publications/2016/05/2836, accessed 23/07/2018.

Scottish Government (2017a), *2018 National Improvement Framework and Improvement Plan for Scottish Education - Achieving Excellence and Equity*, December, http://www.gov.scot/Resource/0052/00528872.pdf, accessed 21/07/2018.

Scottish Government (2017b), *Summary Statistics for Attainment, Leaver Destinations and Healthy Living*, No. 7, 2017 Edition, National Statistics Publication, June.

Scottish Government (2017c), *Science, Technology, Engineering and Mathematics (STEM) Evidence Base*, Scottish Government, October.

Scottish Government (2017d), *Scottish Survey of Literacy and Numeracy 2016 (Literacy)*, May, http://www.gov.scot/Publications/2017/05/7872/341219, accessed 23/07/2018.

Scottish Government (2017e), *National Improvement Framework: Consultation on measuring the attainment gap and milestones towards closing it*, October, http://www.gov.scot/Publications/2017/10/1047/345709, accessed 21/07/2018.

Scottish Government (2017f) Graduate outcomes by University and subject (Longitudinal Educational Outcomes (LEO) data) – Scotland, Scottish Government, June.

Scottish Government (2018), *Evaluation of the Attainment Scotland Fund - interim report (Years 1 and 2)*, http://www.gov.scot/Resource/0053/00532725.pdf, March, accessed 22/07/2018.

Scottish Office (1994), *Higher Still: Opportunity For All*, Edinburgh, HMSO.

Shakeshaft, N. G., Trzaskowskia, M., McMillana, A., Rimfelda, K., Krapohla, E., Haworth, C., Dale, P. S. & Plomin, R. (2013), Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16, *PLoS ONE*, Vol. 8, No. 12: e80341.

Shapira, M. & Priestley, M. (2018), Narrowing the Curriculum? Contemporary trends in provision and attainment in the Scottish Curriculum, *Scottish Educational Review*, Vol. 50, No. 1, pp75-107.

Shavit, Y. & Blossfeld, H. P. (eds) (1993), *Persistent inequality: Changing educational attainment in thirteen countries*, Boulder, CO: Westview Press.

Sosu, E. & Ellis, S. (2014), *Closing the Attainment Gap in Scottish Education*, Joseph Rowntree Foundation.

Stoké, V. (2007), Explaining Educational Decision and Effects of Families' Social Class Position: An Empirical Test of the Breen-Goldthorpe Model of Educational Attainment, *European Sociological Review*, Vol. 23, No. 4, pp505-519.

Sullivan, A. Zimdars, A. & Heath, A. (2010), The Social Structure of the 14-16 Curriculum in England, *International Studies in Sociology of Education*, Vol. 20, No. 1, pp5-21, March.

Tam, T. (2011), Accounting for dynamic selection bias in educational transitions: The Cameron-Heckman latent class estimator and its generalisations, *Research in Social Stratification and Mobility*, Vol. 29, pp287-310.

Tapp, A. L., Maybery, M. T. & Whitehouse, A. J. (2011), Evaluating the twin testosterone transfer hypothesis: a review of the empirical evidence, *Hormones and Behavior*, Vol. 60, No. 5, pp713–722.

Tinklin, T., Croxford, L., Ducklin, A. & Frame, B. (2001), *Gender and Pupil Performance in Scotland's Schools*, Scottish Executive Education Department, September.

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., Gottesman., I. I. (2003), Socioeconomic status modifies heritability of IQ in young children, Psychological Science, Vol. 14, No. 6, pp623–628 in de Zeeuwa, E. L., de Geusa, E. J. C. & Boomsa, D. I. (2015), Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement, *Trends in Neuroscience and Education*, Vol. 4, No. 3, pp69-76, September.

UK Commission for Employment & Skills (2015), *Reviewing the Requirement for High Level STEM Skills*, Evidence Report 94, UKCES, July.

UK Commission for Employment & Skills (2016), *Working Futures 2014-2024 Report*, UKCES.

Van de Beek, C., Thijssen, J. H., Cohen-Kettenis, P. T., van Goozen, S. H., & Buitelaar, J. K. (2004), Relationships between sex hormones assessed in amniotic fluid, and maternal and umbilical cord serum: what is the best source of information to investigate the effects of fetal hormonal exposure? *Hormones and Behavior*, Vol. 46, No. 5, pp663–669 in Gielen, A. C. & Zwiers, E. S. (2018), *Biology and the gender gap in educational performance - The role of prenatal testosterone in test scores*, IZA Discussion Paper No. 11936, November.

Vinkhuyzen, A. A., van der Sluis, S., Maes, H. H., Posthuma, D. (2012), Reconsidering the heritability of intelligence in adulthood: Taking assortative mating and cultural transmission into account, *Behavior Genetics*, Vol. 42, No. 2, pp187–198.

Vuoksimaa, E., Eriksson, C. J., Pulkkinen, L., Rose, R. J. & Kaprio, J. (2010), Decreased prevalence of left-handedness among females with male co-twins: evidence suggesting prenatal testosterone transfer in humans?, *Psychoneuroendocrinology*, Vol. 35, No. 10, pp1462-1472.