Stratification into Field of Study in Higher Education

Natasha Codiroli Mcmaster

UCL

PhD Quantitative Social Research (Education)

Declaration

I hereby declare that, except where explicit attribution is made, the work presented in this thesis is entirely my own. Chapter three is a dual-authored paper, written with Rose Cook from the Institute of Education, UCL. We both contributed equally to this paper, and had joint involvement in the writing of all sections.

Word count (exclusive of appendices, bibliography, tables and front material):
43,538 words

Abstract

This thesis analyses the extent students are stratified into subjects depending on their social background, and the consequences of this in the labour market. I draw on analysis from three longitudinal cohort studies; Next Steps, the 1970 British Cohort Study, and the US study National Longitudinal Survey of Youth (NLSY79). It makes four unique contributions to the literature on educational inequality and subject choice. Firstly, in a joint-authored paper, it offers an overview of the use of intersectionality as a method in quantitative educational research. We make the case that the method should be used more readily in research measuring inequalities in education. Secondly, I empirically test the relationships between students' characteristics, including their social background, ethnicity and gender, and field of study in higher education. I find that parents' level of education is more strongly associated with subject choices than either social class or financial resources, suggesting a preferred focus of future research into stratification by subject. I also find that gender and social background interact in determining choice of degree subject. Thirdly, I go on to explore the psychological mechanisms that may drive differences in subject choices. I find differing relationships between students' personal attitudes and university choices depending on social background. Students from more advantaged backgrounds appeared most likely to choose subjects they enjoyed and thought they were good at. My final chapter compares the relationship between social background and subject choice in the UK and the US. I find that parental education was associated with subject choice for the US cohort, but not the UK cohort. I further test how far these differences explained disparities in earnings in adulthood, and do not find evidence that differences in field of study by background contribute to earnings inequalities in later life.

Impact statement

The research outlined in this thesis forms an addition to the current literature on educational inequalities, field of study, and the intergenerational transmission of (dis)advantage in the UK and the US.

Chapter three aims to contribute to the literature methodologically by giving an overview of quantitative education research using intersectionality as a method. I expect researchers to be able to use this work to inform their own thinking and future research. I hope that this review will increase the reach of the body of work reviewed, and that it will become common practice to test for interactions when considering inequalities both within and across countries, and over time, motivating and contextualising this approach using intersectionality theory. I also hope that the chapter will have impacts on data collection, by stressing the importance of collecting robust data on less represented groups, increasingly harmonising datasets, and linking administrative and survey datasets where possible.

Chapters five, six and seven deal with an issue of high policy relevance; the subjects students study at A level and in higher education. Successive governments have highlighted the need for graduates with skill sets that complement industry demands, as well as the need for greater diversity within particular industries. By outlining the student characteristics associated with subject choice, I hope policy makers will be able to use this information in designing and implementing relevant interventions and engagement events. Chapter six in particular aims to uncover the mechanisms driving disparities in subject choices by students' social background, finding that these differences remain for students with similar enjoyment and perception of ability in science and maths. Organisations that could use this knowledge include the Department for Education, the Department for Business, Innovation and Skills, and think-tanks & charities interested in increasing overall participation as well as diversity in specific fields.

To facilitate greater impact and reach of my research I have published two journal articles based on chapters three and five (Codiroli Mcmaster, 2017; Codiroli Mcmaster & Cook, 2018). I have also submitted my research in chapter six for publication in a peer reviewed academic journal, and am planning to submit research in chapter seven to a suitable academic journal.

I have further presented the findings from my research at the following conferences and events: Department for Quantitative Social Science graduate student seminars, October 2015; Association for Public Policy and Management (APPAM) International Conference, June 2016; Society for Longitudinal and Life course Studies (SLLS), October 2016; European Consortium for Sociological Research (ECSR) Spring School, March 2017; UC Santa Barbara Labor Lunch, December 2017.

Finally, I have written two blog posts for LSE blogs series (Codiroli Mcmaster, 2017a; 2017b) based on my research. The readership of this blog series include (along with researchers and students) policy professionals, think tank researchers and third-sector professionals. The posts are disseminated through social media platforms to reach a wider and more diverse audience.

Table of Contents

List of tables

List of Figures

Acknowledgements

I am very grateful to the Economic and Social Research Council (ESRC), via the Bloomsbury Doctoral Training Centre, for funding the research included in this thesis. I am also grateful to the ESRC for funding my MSc in Policy Analysis and Evaluation at the Institute of Education, which provided vital research training to help me complete my PhD, and without which I would unlikely have been able to continue study full-time.

I would also like to thank the Centre for Longitudinal Studies and the UK data archive for use of the data integral to analysis in chapters five and six, including Next Steps data and the British Cohort Study (BCS70), and the Department of Education for allowing me to access National Pupil Database and higher education Statistics Agency linked data for analysis in chapter five. The National Longitudinal Survey of Youth (NLSY79) was used for analysis in chapter seven, and is made available to researcher from across the world by U.S. Bureau of Labor Statistics, for which I am very grateful.

Finally, it's important to state that all opinions expressed in this thesis are my own, and not those of any funders or data providers outlined. Any mistakes are also my own.

Thanks

I would like to thank my supervisors for their advice and support throughout the writing of this thesis. Lindsey Macmillan for supporting my application for 1+3 funding to study both my MSc and PhD at the Institute of Education. Lindsey was instrumental in supporting the formation of my research questions, and has offered feedback throughout the process with enthusiasm and rigour. Similarly, I would like to thank Alice Sullivan for supervising my thesis, offering invaluable advice, guidance and suggestions for improvement throughout. I have also had two substitute supervisors during Lindsey's maternity leave; George Ploubidis and Jake Anders, who offered differing perspectives to my work and strengthened my final papers.

I was hosted at the University of California, Santa Barbara (UCSB) economics department by Shelly Lundberg whilst writing chapter seven of this thesis. Thanks to Shelly for her sharing her expertise with me throughout this placement, and offering insightful advice on my paper. Thanks also to the economics department for allowing me to present my paper at their Labor lunch seminar series, and for including me in their weekly education reading group.

Thanks to the audiences at the following conferences and events for their thoughtful feedback: Department for Quantitative Social Science graduate student seminars; Association for Public Policy and Management (APPAM) International Conference; Society for Longitudinal and Life course Studies (SLLS); European Consortium for Sociological Research (ECSR) Spring School.

I completed secondments at both the Cabinet Office and the Department for Work and Pensions during my PhD. I would like to thank Research Councils UK for funding this opportunity, which allowed me to develop both my analytical and policy skills.

Thanks to my family, Eliane Codiroli and Chris Mcmaster for fostering my inquisitiveness and contrariness, and to Ruby Codiroli being a ray of sunshine.

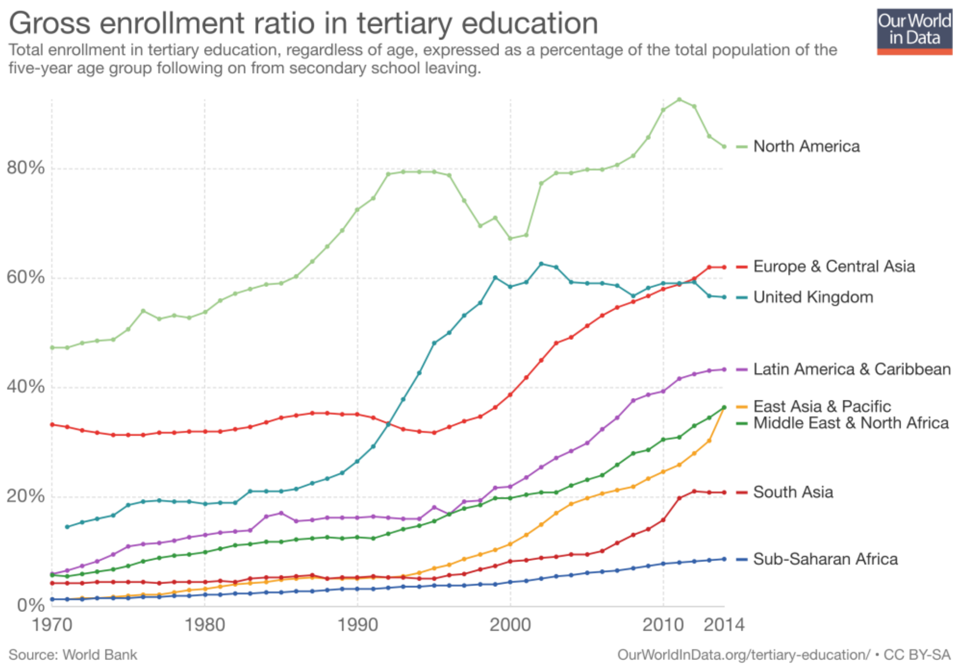
Finally, special thanks goes to Brook Klepper for his emotional support throughout my studies, for proofreading my work and for listening to me practice presentations repeatedly.

Chapter 1: Introduction

The extent that a person's background determines their educational success and participation is an extensive area of research across the social sciences, in education, sociology, economics and psychology. This is unsurprising given the positive outcomes associated with educational achievement, including higher lifetime earnings and better chances of entering more prestigious, rewarding and stable occupations (e.g. Card, 1999; Naylor, Smith, & Telhaj, 2016; Shavit & Muller, 1998; Walker & Zhu, 2008), along with better health (Conti, Heckman, & Urzua, 2010) and wellbeing (Melhuish, 2014). In terms of its associations with social background, many researchers argue not only that the positive benefits of education should be accessible to all regardless of background, but also that education can be a vehicle through which meritocracy is achieved, as people receive rewards based on their achievements rather than background. Often these debates have focused solely on levels of education or attainment, intimating that experiences and rewards are similar within these levels. This thesis adds to this literature by analysing the relationships between social background and field of study within levels of education, and the implications different subject choices have on later occupational outcomes.

Recent demographic trends have led to a shift in thinking about the nature of education, and its relationship with both early experiences and later outcomes. The proportion of people entering any tertiary education is increasing across the world; as shown in figure 1.1, and by 2014 over half of young people in the UK were entering higher education in the five years after leaving compulsory schooling. A consequence of this increase in attendance is that experiences within higher levels are becoming ever more relevant. This is often referred to in the literature as a distinction between vertical stratifications in education, those that occur at different levels, and horizontal ones that occur within the same level (Gerber & Cheung, 2008). There are two main ways students are stratified within educational levels: into institutions of differing quality, prestige, or focus, and into different subject areas. Whilst the relationship between higher educational institution type and social background is an important and vibrant research area (e.g. Boliver, 2013; Jerrim, Chmielewski, & Parker, 2015), this thesis focuses on field of study for a number of reasons.

Figure 1.1: Percentage of the population of the five-year age group following secondary education entering tertiary education each year



Firstly, field of study is more often thought of as a choice students make within a number of possible alternatives, rather than one for which the benefits compared to alternatives are very clear. There is some argument over which universities or higher education institutions are 'better' than others (Boliver, 2015), however there is a general understanding of which institutions would lead to better prospects later in life, as outlined yearly in a number of league tables. It would be expected that, all else held equal, students would choose the universities offering highest potential rewards. There are of course exceptions to this, including where students are constrained by location. Choosing a subject, however, is a somewhat different choice. Whilst attainment is associated with subject studied, a student in any attainment group would have few subjects 'closed off' to them. Instead of simply choosing a subject that will offer highest returns, students choose subjects based on their personal preferences, values and domain specific abilities (Wigfield & Eccles, 2000).

The case of gender difference in science participation is a clear example of the ways people may make choices that are associated with lower economic reward based on preferences, values and perceptions of their ability (Bøe & Henriksen, 2015).

These preferences are often thought to be innate, and recent research suggests genes do indeed influence choice of subjects (Rimfeld, Ayorech, Dale, Kovas, & Plomin, 2016), but so too do early environment and family influences. As Bradley and Charles (2003) point out, cultural norms and stereotypes also influence segregation into subjects. In the absence of full exploration of a subject, people assume they will enjoy subjects in which people with a similar group identity cluster.

Second, field of study is associated with outcomes after university, including probability of finding sustained, graduate employment and higher income returns. The subjects that appear to be most lucrative are STEM subjects, and also Law, Economics and Management (see Chevalier, 2011; Department for Education, 2017; Walker & Zhu, 2011), and these differences in earnings appear to persist when controlling for a persons individual characteristics, for example social background, gender and prior attainment (Sullivan et al., 2018). There are several mechanisms through which different subjects can lead to different returns after graduation. Some subjects may increase human capital directly through superior quality of training. They may also better complement employment demands within societies, leading to increased opportunities and better bargaining power for individuals. Subject choice may alternatively simply act as a signal to employers of higher ability, which would become more important as more people attend university, and it becomes harder to differentiate applicant's based simply on level of education. Regardless, a general consensus is that people should have equal access to these rewards. Where social background is associated with choices, field of study may become a mechanism through which families maintain advantage. The extent that this is the case is empirically explored in chapter seven of this thesis.

Finally, there has been a sustained policy interest in ensuring enough people in the UK are studying subjects and gaining skills that complement the needs of the economy. Concern has been raised that the growing reliance on Science, Technology, Engineering and Mathematics (STEM) skills is not mirrored by increased training in these areas. Successive policy has highlighted a need to focus on up-skilling people in these areas, most recently the Industrial Strategy Green Paper (HM Government, 2017). This skills gap is usually set in a global context, with many other countries seeing a higher proportion of young people study STEM subjects (CBI, 2014; van Langen & Dekkers, 2005). There remain questions about whether the rhetoric around the STEM skills gap is grounded in evidence, and whether STEM university education really prepares people to fill this gap (Smith & White, 2017). Nonetheless, understanding the characteristics associated with subject choice in university has the potential to guide efforts aimed at increasing uptake of specific subjects.

1.1 Thesis outline and contributions

This thesis is comprised of three sole-authored empirical chapters aimed at improving our understanding of the determinates of subject choice in higher education, and their consequences. This adds to the current UK literature looking at inequalities in educational attainment and subject choices at earlier ages, and to international evidence of the consequences of qualitative differences in higher educational experiences. Taken together, the chapters enhance our understanding of the processes involved in making educational choices. In these chapters I argue that differences in subjects studied by a range of student characteristics, particularly social background, are not fully explained by prior educational attainment. Methodologically, I add to the literature by showing that associations between student characteristics and field of study are not simply additive, and demonstrate the importance of accounting for multiple characteristics and contextual differences (for example, the different educational systems in the UK and the US). This is further elucidated in a joint-authored chapter on the contribution of intersectionality to research in this area, also included in this thesis.

In chapter two I give an overview of the literature into educational inequalities more broadly, giving some perspective to how our thinking about inequalities within education has developed, and context of the state of educational inequalities in the UK.

Chapter three offers an overview of the use of intersectionally as a method in quantitative research addressing educational inequalities. This dual-authored chapter argues that, where appropriate, a consideration of the impacts of belonging to multiple groups should be more often considered in this research area. I, with my co-author Rose Cook, show that there is an emerging literature suggesting that experiences and outcomes of young people differ depending on the combination of characteristics and group memberships. These characteristics may reflect a person's own identity, how their behaviour may change given this identity (for example, higher aspirations or interests), or differences in treatment by others based on perceived group membership. A version of this chapter has been published by the British Educational Research Association journal, Review of Education (Codiroli Mcmaster & Cook, 2018).

In chapter four I outline the data used in the empirical chapters in this thesis, and rationalise the use of longitudinal surveys over other data sources. I compare the data with administrative sources where possible, showing that the data appear representative of the population of interest according to some key demographic characteristics. I also discuss important methodological decisions, including the measurement of family background and the treatment of missing data.

Chapter five draws on ideas from chapter three by outlining the current state of disparities in the subjects young people study in post-compulsory education. This chapter focuses on choice of STEM subjects. This group of subjects that have received the most policy and research interest over the years due to high levels of gender segregation and the perception of increasing importance they have for the economy. I examine associations between subject studied and gender, ethnicity and family background. I not only ask whether family background is associated with field of study, but *which* family background characteristics drive observed disparities. Furthermore, I take an intersectional approach, describing differences in choices by multiple group identities. I find that parents' level of education is more strongly associated with subject choices than either social class or financial resources, suggesting a preferred focus of future research into stratification by subject. I also find that gender and social background interact in determining choice of degree subject. A version of this chapter was published in the British Educational Research Journal in 2017 (Codiroli Mcmaster, 2017), with an earlier version published in the peer reviewed Centre for Longitudinal Studies working paper series in 2015 (Codiroli, 2015). I have also written a blog post based on this chapter for the London School of Economics (LSE) British Politics and Policy blog series (Codiroli Mcmaster, 2017a).

In chapter six I explore the psychological mechanisms that may drive differences in subject choices. This follows literature suggesting students' own perception of ability and enjoyment of subjects impact subject choices, and that these attributes differ by students' characteristics. This literature highlights these processes as key mechanisms behind gender disparities in field of study, and I test whether this is also the case for social background disparities. I also consider whether the processes behind subject choices differ by student characteristics by asking whether students are less or more likely to study subjects they are passionate about, or think they are good at, depending on their relative advantage. My first hypothesis was not supported; students' social background was still associated with choices when accounting for differences in enjoyment and perception of ability. There were, however, differing relationships between students' personal attitudes and university choices. Students from more advantaged backgrounds appeared more likely to choose subjects they enjoyed and thought they were good at. The British Journal of Sociology in Education has accepted this paper for publication.

Chapter seven aims to measure the implications of differences in subject choices by background in both the UK and the US on later outcomes. This chapter is unique in exploring associations between social background and field of study in two Anglophone countries, and in focusing specifically on subject studied as a possible driver of income disparities in later life outcomes. It also utilises policy relevant subject categorisations that take account of gender differences in field of study, and the impact of background on educational choices. I discuss structural differences between the two countries that may give rise to differences in results. Overall, I do not find evidence that differences in field of study by background contribute to income inequalities in later life.

In my final chapter I offer a summary of these three empirical chapters and the review chapter, outlining their findings and contribution to the literature. I then discuss current changes in policy and the implications these may have for findings. I finally offer some suggestions for future research and ways to expand our knowledge about field of study even further.

Chapter 2: Overview of the literature on educational inequalities

2.1 Social inequalities in education

In understanding the causes and consequences of social stratification into subject areas, it's important first to have a broader understanding of the ways students are stratified in education more generally. Largely, researchers interested in inequalities in education have focused on differences in achievement in standardised tests, the ways these differences develop over time, and their consequences on later outcomes. This has been referred to as Inequality in Educational Opportunities (IEO), and the effects of IEO include lower chances of entering higher levels of education or securing professional employment in adulthood. Not only have large differences in educational attainment by social background been recorded, but this is a global phenomenon, which research suggests has either not changed since the 1960's (Shavit & Blossfeld, 1993) or is reducing at a slow rate (Blanden & Macmillan, 2016; Breen & Jonsson, 2005; Breen, Luijkx, Müller, & Pollak, 2009, 2010).

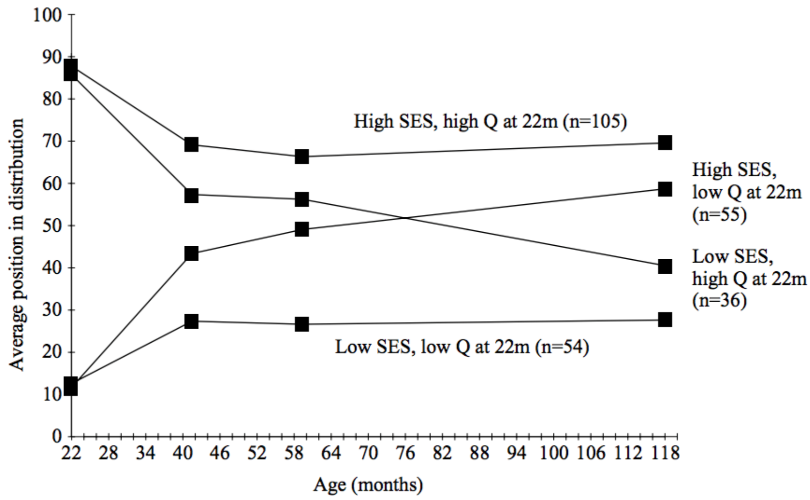
The raw statistics on attainment differences by children's levels of advantage in the UK are stark. A recent report by the Education Policy Institute shed light on the extent of these disparities, using the National Pupil Database (NPD) records of exam results including all school students in England. Students eligible for Pupil Premium, additional funding that is given to schools for each disadvantaged student they admit, were just under 2 years behind their peers in terms of attainment by the end of secondary school. They also find that, whilst these differences begin relatively small, they increase each year, resulting in a cumulative effect of disadvantage (Andrews, J., Robinson, D., & Hutchinson, 2017).

If we focus purely on the raw statistics, questions remain about whether these associations really are indicative of a society in which people fall behind because of their background, or whether they simply reflect the genetic differences in ability between children (e.g. Gottfredson, 2004). Researchers studying the genetics of intelligence have however only been able to explain around half of the variation in exam scores, with the rest likely attributable to environmental differences (Chipuer, Rovine, & Plomin, 1990; Johnson, McGue, & Iacono, 2006; Loehlin, 1989).

The fact that disparities increase over an individuals life-course can either be attributed to the increased exposure to different environments over time, the effect of persistent disadvantage and poverty, or gene-environment interactions (where children may have a predisposition to particular traits and are exposed to environmental triggers) (see Rutter, 2006).

In testing how children's background and attainment are associated over the life-course, researchers have turned to longitudinal cohort studies, which follow thousands of children from birth into adulthood and included rich information on cognitive ability and social background. Leon Feinstein analysed the 1970 British Birth Cohort and found that not only did differences between children grow over time, but children from advantaged families who initially scored poorly in academic tests appeared to overtake their less advantaged, higher attaining, peers in a relatively short period of time (see Figure 2.2) (Feinstein, 2003). This seminal paper has influenced policy arguments on how far we should intervene in maintaining fairness in schools, with the suggestion being that there is hope in halting the reduction in attainment of poorer students over time if we intervene early. Indeed, evidence from the US suggests that some interventions starting very early in a child's life do help to close gaps in achievement (Cunha & Heckman, 2009; Heckman, 2006). Whilst Feinstein's findings may be less stark than previously thought due to 'regression to the mean,' where children who initially receive very high scores receive scores closer to the average in later testing (Jerrim & Vignoles, 2013), the effects largely remain when accounting for this, and subsequent research using more recent cohorts of children has found similar results (Dickerson & Popli, 2016). By 16 years old (a pivotal age in deciding whether to continue with education, and which educational pathways to take) students from less advantaged backgrounds are much less likely to achieve the common benchmark of educational success; 5 GCSEs at grades A*-C (Strand, 2014a).

Figure 2.2: Average rank of test scores at 22, 42, 60, and 120 months, by SES of parents and early rank position, from (Feinstein, 2003)



Students are not only disadvantaged in later life through lower test scores, but also through the educational choices they make regardless of attainment. Quantitative research into the ways student background influences choices was conducted as early as the 1930's in Sweden (Boalt & Janson, 1953), however the majority of studies continued to either focus on attainment differences, or see disparities in educational transitions as caused purely by ability differences. This early work was revisited by Girard and Bastide's (1963), where attainment differences were labelled 'primary effects' and differences in educational choices were labelled 'secondary effects'. Boudon further developed this concept, arguing that secondary effects of social background on education arise from the fact that there are different benefits, and costs, to remaining in education depending on family resources (Boudon, 1974).

Given the long history of the concept, surprisingly little attention had been paid to secondary effects until relatively recently (Jackson, Erikson, Goldthorpe, & Yaish, 2007). Despite this, there are now a number of studies giving strong evidence that students' choices to attend higher levels of education are influenced by their background, even when accounting for achievement differences. This inequality has received considerable public interest, with the creation of a number of charities and public bodies aiming to increase participation of bright but disadvantaged students in higher education (e.g. the Sutton Trust, the Access Project, the Office for Fair Access, etc). In 2016 just 16% of the students who claimed Free School Meals attended university in the year after secondary school, compared with just under 33% of students who did not (UCAS, 2016). Blanden and Gregg (2004), using two longitudinal cohort studies, found a positive association between family income and attendance at university in the UK even when controlling for prior achievement. Unlike disparities in test scores, there is no evidence that secondary effects of social backgrounds are decreasing over time. This relationship appeared to be increasing for people born in 1958 and 1970 (Blanden, Gregg, & Machin, 2005), people from more advantaged backgrounds born in 1970 were even more likely to go to university than their less advantage peers, compared to people born in 1958.

2.2 Social inequalities beyond education - the Direct Effect of Social Origin (DESO)

The effect of social background that remains when accounting for education is the Direct Effect of Social Origin. Whilst the majority of studies point to some remaining association (Bernardi & Ballarino, 2016; Gregg, Jonsson, Macmillan, & Mood, 2017; Gugushvili, Bukodi, & Goldthorpe, 2017), there is a lack of consensus in the literature about whether DESO remains when accounting for education. Results differ depending both on the way education is measured, and the way achievement in adulthood is measured (i.e. income, or social class). The direct relationship between parental characteristics and their child's income is substantial in both the UK and US. Some research even suggests that the association is larger in the UK than the US, and that 'The American Dream,' although still elusive, is indeed more attainable in the US (Bernardi & Ballarino, 2016). In the UK, even when less advantaged people do reach a similar social position to their more advantaged peers, they face a considerable earnings penalty (Friedman, Laurison, & Macmillan, 2017; Laurison & Friedman, 2016). Gugushvili and colleagues (2017), along with finding a strong independent relationship between background and social class when additionally accounting for education completed outside 'non-traditional' ages, found that part of this association could partly be accounted for by differences in individuals' cognitive ability and locus of control.

Few studies have explicitly taken account of horizontal stratification, or subject choices, which leads to an overestimation of the DESO. Those that do, show a reduction or even disappearance of DESO (Belfield et al., 2018; Gregg et al., 2017; Jacob, Klein, & Iannelli, 2015). The former study linked administrative Higher Education (HE) and tax records, finding that the association between parents and their child's income falls, but is not entirely explained, when accounting for subject choice and institution. This research was based on students' entry to university, and it remains possible that the difference in earnings could in part be down to different university drop-out rates by background (Crawford, 2014). Gregg and colleagues (2017) compared the intergenerational transmission of income in the UK, US and Sweden, finding that when controlling for education, including field of study, the relationship remains for men. In contrast, Sullivan et al., (2017) find that when controlling a complete picture of educational experience there is little relationship between social origin and gaining access to the top social classes. Chapter seven of this thesis explicitly addresses the extent that differences in field of subject may account for some of this 'unexplained' effect of social background on earnings.

2.3 Field of study and social background

All empirical chapters in this thesis are concerned with the relationship between social background and field of study at university. There is an emerging literature now looking beyond inequalities in achievement within education and progress to different levels of education, instead focusing on differences in choices within levels of education. These distinctive elements within education are often referred to in the literature as 'horizontal' inequalities (Charles & Bradley, 2002; Gerber & Cheung, 2008). Largely, this research has been conducted using samples in mainland Europe and North America, and less research has been conducted in the UK. This thesis fills this gap in the literature by explicitly analysing the associations between social background and field of study in England and the UK, using two nationally representative cohorts.

In one of the earlier studies considering horizontal inequalities, Van de Werfhorst and Luijkx (2010) studied the subject choices of Dutch men attending university. They found strong similarities in the field of men's university choices and father's occupation. They analysed a more fine-grained measure of occupation beyond the broad social class categories, arguing that social stratification occurs at a more domain specific level. For example, children of medics would be more interested in medicine, or related fields, than children of Engineers, even if they fall within the same social class. In a later study, complementing findings in chapter five of this thesis, Van de Werfhorst (2017) finds that gender and social background interact in determining choices of Dutch students. Both women and men were less likely to choose subjects that were 'gender typical' if they had more educated parents. In Italy, Triventi, Vergolini and Zanini (2017), using repeated cross sectional data, similarly found that people from more privileged backgrounds were more likely to study subjects associated with higher incomes, and this relationship remains when accounting for attainment in school.

2.4 Returns to field of study in university

Where students' background is associated with choices, different returns to subjects could increase the intergenerational transmission of advantage (something directly tested in chapter seven). There has been much research into *which* subjects confer higher returns. Notably, Ian Walker and Yu Zhu have published a number of papers looking at the relationship between subject studied and earnings in the UK (Walker & Zhu, 2001, 2008, 2011, 2013). They find that law, economics and management (and related) subjects confer the highest earnings returns, and Other Social Sciences, Arts and Humanities (OSSAH) the lowest returns. These findings remained robust when different data sources were used, and when the focus was either point in time or lifetime earnings. More recently, the Department for Education have linked university data with tax records, to estimate the earnings differences across all subjects. This has allowed for a more in-depth analysis of the average returns and distribution of returns by subject, without restrictions based on sample sizes. The findings from this dataset are broadly similar to those of Walker and Zhu, with medicine, veterinary sciences and economics associated with highest earnings, and creative arts and design associated with lowest earnings. They also identify subjects with large heterogeneity of returns, for example law, business and administrative subjects and economics, where *some* students go on to obtain extremely high paying jobs (Department for Education, 2017). This research has even led to suggestions that, due to their lower expected returns, arts and humanities students should pay lower fees (Shipman & Griffiths, 2018).

A question that remains following this research is whether these subjects *cause* higher earnings returns, or whether they simply are more likely to admit students who would have earned more whatever they studied. This may occur, for example, through subjects admitting students with higher ability, or students from more privileged backgrounds. Richard Blundell and colleagues (2000) used the National Child Development Study, a longitudinal birth cohort of people born 1958, and found that differences in returns by subject persisted when matching people based on cognitive test scores, qualifications and social background, suggesting that the subjects themselves impact earnings. Similar results have been shown using the 1970s British Cohort Study (Bratti, Naylor, & Smith, 2008), and law, economics and management studies have been shown to have the strongest association with entering the highest paid professions when controlling for ability, background and previous qualifications (Sullivan et al., 2018). There is a gap in the literature looking at more recent returns to subjects, and this could have changed substantially following the expansion of the university system. Whilst Belfield et al (2018) show that differences in earning by subject persist when controlling for background using more contemporary data, their use of administrative data did not allow them to control for ability.

Researchers have also tested how far differences in field of study impact differences in occupational outcomes by social background. Triventi (2013) used data from Germany, Norway, Italy and Spain to not only test the hypothesis that field of study mediates intergenerational mobility, but also whether this differs in countries with very different education systems. In Norway, Italy and Spain, more advantaged young people choose more lucrative subjects, and this did indeed help explain why parents and their children had similar occupational outcomes. Jacob, Klein and Iannelli (2015) also find that, in both the UK and the US, field of study partially mediates the relationship between parent's education and their children's occupational class, based on cross-sectional survey data. Chapter seven in this thesis, however, using a representative longitudinal cohort, finds that whilst the association between parents' education and people's incomes at 42 years old reduces when controlling for degree attainment, additionally controlling for subject studied did not add any explanatory value. That is, for the cohort born in 1970 in the UK, stratification into field of study does not seem to explain intergenerational transmission of advantage.

This has also been a vibrant area of study in the US. Davies and Guppy (1997) analysed NLSY79 data, a representative longitudinal study of young people in the US, and ranked subjects by average monthly income in adulthood. They found that social disparities in subject studied only arose for students attending more prestigious universities, with higher SES students more likely to study subjects leading to higher income returns in the future. By focusing on potential income differences between subjects, this did not analyse social differences in more nuanced dimensions of subject choices, for example perceived difficulty or the extent to which they were 'traditionally academic'. Goyette and Mullen (2006), using data from the National Educational Longitudinal Study (NELS) and the Baccalaureate and Beyond Longitudinal Study (B&B), grouped subjects based on vocational focus. They found that lower SES students were most likely to study vocational courses, and least likely to study liberal arts and sciences. This analysis controlled for differences in student achievement and the characteristics of the university attended. Both studies created a composite score of social background including some combination of parents' occupational class, education and/or income.

Interested in the mechanisms behind stratification, Moakler and Kim (2014) noted that parents who had higher education levels and incomes were also more likely to be working in STEM occupations. They analysed survey responses of freshmen and found that having parents working in a STEM field, but not more traditional social background measures, increased students probability of choosing STEM at university. Ma (2009) using NELS 1988-1994 data found that students from less advantaged backgrounds were *more* likely to study fields with higher economic returns. She then found that this effect only held for women, men did not seem to be influenced by family background in course choices. Similarly, Leppel, Williams, & Waldauer (2001) show that women whose fathers work in a professional occupation are less likely to major in Business than women whose fathers work in less prestigious occupations. Chapter seven in this thesis uses US data to help further understand how the relationship between subject choice and social background impact income.

This thesis focuses primarily on UK students. One of the first papers in the UK to explicitly test this relationship between background and field of study in a nationally representative sample found that students from higher social class backgrounds were more likely to study 'prestigious' subjects at university, including medicine and law (Herman G Van De Werfhorst, Sullivan, & Cheung, 2003). This sample included individuals born in 1958 who would have attended university in a very different educational context, most notably with far fewer students entering university at all. It is likely that with fewer students getting any degree, the subject and type of degree would have less bearing on later occupational outcomes. This thesis adds to the knowledge about the relationship between background and field of study by using contemporary longitudinal datasets, including a cohort of people born in the UK in 1970, and young people living in England born between 1989 and 1990.

Chapter 3: The contribution of intersectionality to quantitative research into educational inequalities

3.1 Introduction

Inequalities in education are one of the most enduring social problems in contemporary societies and have been examined extensively in social science research. People from the most privileged backgrounds dominate educational opportunities, and this is related to the inter-generational transmission of socio-economic position (Breen & Jonsson, 2005; Breen, 2009, 2010; Ishida et al., 1995). Inequalities in educational outcomes contribute to differences in civic participation (Marien et al., 2010), wellbeing (Melhuish, 2014), earnings (Checchi and Van de Werfhorst 2017) and health (Conti et al., 2010). These inequalities also have implications for countries' economic prosperity (Hanushek & Woessmann, 2008). A myriad of policy proposals and social programmes have been initiated aiming to tackle educational inequality, yet there appear to be no straightforward solutions, and research on its patterns, trends and mechanisms is ongoing.

An obvious first step to tackling educational inequality is defining the problem adequately. In political and public discourse, 'educational inequality' is often framed in simplistic, vague terms, referring to individuals who are more or less privileged with respect to education. However, this description obscures a highly complex reality. Multiple aspects of advantage and disadvantage, both separately and in combination, influence educational outcomes. This can include socio-economic background, gender, and ethnic background, among other influences. In this chapter, we argue that the concept of 'intersectionality', derived from feminist theory, is a useful lens through which to view these interlocking disparities in education, and with which to better define and understand the problem of educational inequality. Noting that the concept has been used extensively and effectively in qualitative research into educational inequality, we discuss the possible *contributions of the intersectionality approach to quantitative research on (vertical and horizontal) educational inequalities (attainment and subject choice)*. Applying an intersectional approach has already expanded thinking about educational inequalities, yet there are challenges to overcome if it is to be fully embraced by quantitative educational researchers. In particular, quantitative researchers need to acknowledge that intersectional inequalities have evolved over time as a result of specific historical and contextual conditions

'Educational inequalities' are systematic variations between individuals based on their social group membership (gender, ethnicity, social class), including access to education, experiences, outcomes and returns to education (Jacobs, 1996; Gross et al., 2016a). The chapter focuses on educational inequalities across two important educational outcomes: attainment and subject choices. We thereby distinguish between 'vertical' inequalities, which separate individuals in a hierarchical fashion according to the amount or level of education completed, and 'horizontal' inequalities, which relate to differences within a given level of education (for example, degree subjects) (Gerber & Cheung, 2008). The reason for considering both vertical and horizontal inequalities is that both are associated with life chances. Across the world, grades and qualifications strongly influence individuals' opportunities in the labour market, leading to higher earnings, higher chances of entering more prestigious occupations and higher employment rates (Barone & Van De Werfhorst, 2011; Sullivan et al., 2017), as well as structuring individuals' lives in a range of other important ways (see Pallas, 2000). However, it is becoming clear that subject choices also shape these outcomes. For example, choosing the 'right' subject can determine income returns to a given level of education (Britton et al., 2016; van de Werfhorst et al., 2003).

The present article focuses on quantitative educational research. The concept of intersectionality has historically been much more widely used in qualitative educational research, where it has been a pivotal concept for theorising the experience of inequality and discrimination (for example, see Gillborn, 2015; Gillborn et al., 2012). However, owing to a perception that feminist-informed theory and quantitative methods are incompatible (Scott, 2010), the concept of intersectionality has been less commonly deployed in quantitative educational research. Therefore, to our knowledge there is no review covering intersectional inequalities in education from a quantitative perspective (although see Gross et al. 2016b for an overview of qualitative, quantitative and mixed methods approaches). The aim of this chapter is to show that there is in fact a close fit between the concept of intersectionality and certain quantitative research techniques and to advocate for a wider, more explicit use of this concept in quantitative educational research.

The research questions addressed in this chapter are:

- How can an intersectional perspective be applied to the quantitative study of inequalities in educational outcomes?
- What are the main findings of research considering the intersections between socio-economic background, gender and ethnicity? How can these results contribute to an intersectional understanding of educational inequality?
- What are the methodological challenges associated with using the concept of intersectionality in quantitative educational research?

The first part of the chapter outlines the concept of intersectionality and why it is relevant for studying inequalities in education. We then describe the methodological techniques typically used by quantitative researchers when assessing complex inequalities in education. The third section reviews quantitative educational research that has employed the concept of intersectionality, either explicitly or implicitly, to studying these complex inequalities. We highlight the contributions of these studies to knowledge on educational inequalities, while engaging with critiques that this type of research is not fully 'intersectional'. We further describe the methodological challenges involved in applying intersectionality to quantitative research on educational inequalities and suggest methodological innovations that would facilitate its use to greater effect. Finally, the chapter summarizes the points raised and concludes with several recommendations for future research.

3.2 Origins of intersectionality

'Intersectionality' refers to the idea that social categories, principally those that involve inequality or power, such as gender, race or ethnicity, and social background, are almost always permeated by one another. One's specific location, at the interface between these categories, determines one's experience of the world. The term is often attributed to the American legal scholar Kimberlé Crenshaw, who, in two influential articles (Crenshaw, 1989, 1991), drew attention to the unique disadvantages faced by African American women. Crenshaw's observations became, for researchers and activists, a way to frame complex forms of discrimination and to draw attention to 'interlocking systems' of inequality (Hill Collins, 2002). The theoretical advances of Crenshaw and others built upon an existing critique of the second-wave feminist movement as being dominated by the concerns of relatively advantaged white, middle class women, overlooking the experiences of women facing additional disadvantages related to ethnicity or social status. While intersectionality is most closely associated with gender studies (Lutz et al., 2016), it is now gaining attention across the social sciences. This has led to in-depth reviews of how the concept can be applied in health research (Hankivsky, 2011), sociology (Choo & Ferree, 2010), family studies (Few-Demo, 2014), and psychology (Else-Quest & Hyde, 2016).

3.3 Intersectionality and educational inequalities

Notwithstanding its increasing popularity as a conceptual tool for social science research, the definitive meaning of the term 'intersectionality' is somewhat elusive, and it has been used in various ways (Davis, 2008). It is sometimes used more broadly to describe a perspective on inequality, which emphasises its multi-dimensionality and contextuality, and sometimes refers to more specific research techniques. McCall (2005) summarizes the different uses of intersectionality in social science: to deconstruct social categories such as gender, ethnicity and class (termed 'anti-categorical complexity'); to analyse differences and similarities *within* social categories ('intra-categorical complexity') or to focus on multiple, intersecting inequalities between social categories ('inter-categorical complexity'). All three variants have been deployed to address the issue of educational inequality (Gross et al. 2016b). Studies discussed in this chapter mainly use the 'inter-categorical complexity' approach, since this is the most obviously applicable to quantitative methods (Gross et al. 2016b). However, we will go on to argue that 'intra-categorical complexity' can also be addressed to some extent using quantitative methods.

We concentrate on social background, gender and ethnic disparities, as these are the best-researched and most pervasive forms of inequality in education (see Buchmann et al., 2008; Heath et al., 2008; Marks, 2005a, 2005b; Shavit & Blossfeld, 1993; Gross et al., 2016a). Social background inequalities (also referred to as socio-economic status (SES) inequalities) are defined as differences in educational outcomes between those with more financial, cultural and/or family resources, and those with fewer such resources. Gender inequalities are differences in educational outcomes between males and females[1]. This is a complex issue, since both males and females can be disadvantaged in different areas and stages of education (Buchmann et al., 2008). Research on ethnic inequalities in education often focuses on the disadvantages faced by ethnic minorities (Heath et al., 2008). However, as we will describe, some studies have identified majority groups as being more vulnerable to certain disadvantages.

An intersectional, 'inter-categorical' perspective on inequality recognises that it is not sufficient to focus on ethnic, gender or social background disparities alone; instead, these multiple identities combine to produce 'complex inequality' (McCall, 2001). A focus on 'complex inequality' seeks to correct the idea that different types of (dis)advantages stand alone or are the same for every individual who experiences them (Ferree & Hall, 1996). Ethnic, gender or social background inequalities in educational outcomes may even stem from similar sources. For example, social norms around gender and education, which may inform gender differences in subject choice, can be linked both to gender ideology and to patriarchal control of economic and political resources, which is inherently linked to class inequality (Browne & Misra, 2003) and the exclusionary practices of powerful, privileged groups (Hill Collins, 2002; Weber, 2001). Thinking 'intersectionally' about inequality in education therefore requires a fundamental shift to thinking about a person's whole set of characteristics and circumstances, and how this relates to systems of power and discrimination within and beyond education.

3.4 How can an intersectional perspective on educational inequality be used in quantitative research?

As mentioned in the introduction, the concept of intersectionality has historically been much more widely used in qualitative than quantitative educational research. Gross et al. (2016b) suggest that this is because qualitative research is better suited to analysing complexity and the everyday experience of inequalities. Other authors

[1] The majority of quantitative studies employ a binary definition of gender and this is reflected in our article. As more fluid gender identities are becoming recognized, incorporating more diverse categories would enhance quantitative data collection.

have suggested that, due to its focus on assigning individuals to pre-defined categories, quantitative research is incompatible with an intersectional perspective (Spierings, 2012). From quantitative researchers, there has been concern about the use of small samples (lacking external validity) in research aiming to capture wider social processes surrounding inequalities (Scott, 2010). However, despite these tensions, we suggest that the most important aspects of an intersectional perspective on inequality - multi-dimensionality and contextuality – are amenable to a quantitative research approach (Scott, 2010). Moreover, with innovations in data collection and moves towards inter-disciplinarity and multi-method research, quantitative research on inequality should increasingly be embracing intersectional theory.

Quantitative research into intersectional inequalities mainly relies on secondary data analysis, using large-scale survey or administrative data. For example, in the UK, researchers have used longitudinal data sources such as the Millennium Cohort Study (MCS), which contains detailed information on family background, early development and educational attainment for a representative sample of 19,000 children born in the UK in 2000-2001. Another key source is the Longitudinal Survey of Young People in England (LSYPE, now known as 'Next Steps'), which has been linked with administrative data on educational attainment routinely collected by the UK government. Administrative datasets, such as the National Pupil Database (NPD), and the Higher Education Statistics Agency (HESA) data on university students, are also rich sources in their own right. Many other European countries have detailed administrative records linking education and outcomes, and there are several widely available survey datasets in the US, including the National Longitudinal Survey of Youth (NLSY) 1979 and 1997, and the National Education Longitudinal Study (NELS 1988).

The main analytical techniques used to study 'inter-categorical' intersectional inequalities in education are interaction effects and sub-group differences. While these are not complex methods, they have the potential to deepen and contextualise more conventional analysis of inequality. First, one must identify raw differences between groups, such as differences in mean scores, or proportions of people selecting particular subjects. Researchers can use regression modelling to identify unique associations between, for example, gender and the likelihood of selecting a Science, Technology, Engineering and Maths (STEM) subject, while controlling for other factors that might affect the outcome. Research that stops here assumes that associations between characteristics and outcomes are purely additive. Using the previous example, an additive interpretation would be that the lower likelihood of women studying STEM is independent from the lower likelihood of socially disadvantaged students studying STEM.

In contrast, an intersectional approach to analysing inequalities acknowledges that characteristics like gender and social background interact statistically. For example, the impact of growing up in a low-income family on STEM choice may differ by depending on a young person's gender. To identify these interactions, researchers can run regression analyses for young men and women separately, to see whether social (dis)advantage influences subject choice in similar or different ways for each gender (Harnois, 2013). This can be done by comparing the sign or size of coefficients and is known as a sub-group approach or split-sample regression. An alternative is to add an interaction term to the regression model. A statistical interaction is present when the effect of an independent variable (such as social background) on a dependent variable (such as STEM choice) differs depending on the value of a third variable (such as gender) (Jaccard, 2001). Interactions are usually set up in terms of a 'focal' and a 'moderator' variable. In our example, the focal variable is social background, and we want to see whether its association with STEM choice is moderated by gender.

3.5 Prior research applying intersectionality to the quantitative study of educational inequality

In this section we give an overview of the main applications of an 'inter-categorical', intersectional approach within quantitative research on educational inequality, concentrating on attainment and subject choice. We suggest that, whether or not they explicitly use intersectionality theory, these studies contribute to an intersectional understanding of educational inequality. We also discuss research that attempts to explain these intersectional disparities by considering aspirations, stereotyping and discrimination, and contextual factors (such as location and policy). Reflecting the approach of the majority of studies discussed, we structure this section with specific axes of inequality in mind (e.g. gender and social background; gender and ethnicity). Studies were identified using academic databases and search engines, focusing on research published in the English language[2].

3.5.1 Attainment inequalities

In terms of inequalities in attainment at school, the intersection between social background and gender has been a prominent theme; across the world, boys and young men appear most susceptible to the effects of disadvantage on educational attainment (OECD, 2015). The vulnerability of boys with less educated parents, from low-income backgrounds and/or with absent fathers has been identified as early as age three. For example, Mensah and Kiernan (2010) show that boys' family and local area characteristics disproportionately affect early test scores compared to girls from similar backgrounds. Entwisle et al. (2007) show that the early reading scores of boys who receive meal subsidies, a measure of family financial disadvantage, are lower than those of girls in similar circumstances. Among children not receiving meal subsidies, there is little gender difference in reading scores. These findings suggest that there is an interaction between social background and gender in relation to educational attainment from the earliest stages.

Ethnicity also interacts with both gender and social background in determining academic outcomes. Using nationally representative UK data from Next Steps, Strand (2014a) shows that the socio-economic gradient (the difference in attainment between students from low SES and high SES backgrounds) is particularly large for white boys, compared to other ethnic groups, and compared to girls. At age 16, disadvantaged white and black Caribbean boys are the worst performing groups (Strand, 2014a). The attainment of white, low SES boys declines throughout secondary education at a faster rate than girls from similar backgrounds, and compared to low SES boys from ethnic minority groups (Burgess et al., 2009). In contrast, advantaged white students do disproportionality well compared to advantaged students from other ethnic groups (except for Indian students). Similar patterns have been found in the Netherlands (Dekkers et al., 2000).

[2] Notable studies have also been published in other languages (e.g. Gottburgsen & Gross, 2012) but have not been consulted for this article.

These findings suggest that previous studies showing that social background is related to attainment (e.g. Breen & Jonsson, 2005; Goldthorpe, 1996) may have overlooked important facets of educational inequality by not considering intersections with gender or ethnicity. Findings showing different outcomes for less advantaged students by gender and ethnicity helps to demonstrate a key aspect of intersectionality theory: that not everyone experiences disadvantage in the same way. An intersectional framing of educational inequality also directs our attention to differences among more advantaged students, showing that low attainment among less advantaged white boys should not be the only cause for concern. Among higher SES UK students, Pakistani, black African and Bangladeshi boys, and black Caribbean boys and girls are achieving poor academic results compared to their white counterparts (Strand, 2014b). Similarly, a study in the United States by Bécares and Priest (2015) found that both racial and gender differences in academic outcomes were most pronounced among higher SES students. This shows that the educational benefits of being socially advantaged are not necessarily evenly distributed across ethnic groups, or between males and females. As well as being particularly vulnerable to the effects of disadvantage, white boys seem to derive disproportionate educational benefits from more advantageous social origins.

3.5.2 Subject choice

As noted in the introduction, attainment differences are not the only way inequalities in education are expressed. Students also choose to study different subjects depending on their gender, social background and ethnicity. Research on inequalities in subject choice tends to focus primarily on gender differences in STEM participation (see Boaler et al., 2011). In the UK, while STEM attainment for girls and boys has converged over time, boys remain much more likely to study non-compulsory STEM subjects, particularly Maths, Physics, Chemistry and Engineering (Smith, 2011). Ethnic and social background differences in STEM participation are less well researched. However, white and black Caribbean students have the lowest representation of all ethnic groups in STEM courses, while south Asian students are the most highly represented (Boaler et al., 2011; Equality Challenge Unit, 2015; Jones & Elias, 2005). There is also an emerging literature showing how students' social background is associated with STEM study

(Campaign for Science and Engineering, 2014; Codiroli McMaster, 2017; Gorard & See, 2009). Research taking an intersectional approach has the potential to shed light on how these factors work together in determining subject choice.

For socially advantaged young people, gender appears to have less of an influence on subject choice. However, the nature of this relationship varies across countries. The US literature consistently shows that the effect of family background on subject choices is more pronounced for women than for men (Leppel et al., 2001; Ma, 2009; Trusty et al., 2000). Ma (2009) finds that, while family socioeconomic status and gender both have independent effects on the choice to study technical, life and health sciences, and business at university (compared to social sciences), the effect of social background appears stronger for young women. Compared to more advantaged women, women from disadvantaged families were more likely to study subjects associated with more lucrative careers. For young men, social background had little influence on choices. In the UK, Codiroli Mcmaster (2017) also found that the association between social background and subject choice was stronger for young women than for young men, but in a different direction. Less advantaged women were more likely to study social sciences, law, and business (instead of STEM) compared to their more advantaged peers. Van de Werfhorst (2017) found similar patterns in the Netherlands; young men and women from less advantaged backgrounds were more likely to choose 'gender typical' subjects. The reasons for these cross-country differences are not yet clear, and more research is needed to better understand the influence of national context. What is certain, however, is that it is important not to assume results will be similar across contexts, as the main driver of inequalities by characteristics such as gender and social background are not the characteristics themselves, but the systems of power that create and sustain them.

Research also points to differing associations between ethnicity and subject choice for young men and women. In a US study, Catsambis (1994) found that the over-representation of boys in mathematics courses in Middle and High school was strongest for Latin American students and smallest amongst African American students. Codiroli Mcmaster (2017) also found some evidence of an interaction between gender and ethnicity in university subject choice in the UK. While, in general, black African students are more likely than white students to choose STEM over arts and humanities, this disparity is much more pronounced for young women than for young men. However, Ma's (2009) study on subject choice in US universities did not find any interactions between ethnicity and social background. As with gender, it is highly likely that the experience of being from an ethnic minority background differs hugely depending on context. Moreover, the ethnic groups under consideration also vary widely across contexts. Ethnic minority groups studied in the US (usually black, Latin American, or 'other') will often be very different from those studied in the UK (usually a much broader categorisation).

3.5.3 Stereotypes and identification with STEM

Explanations for gender differences in subject choice have typically focused on social norms about which subjects are appropriate for each gender and how these are internalised throughout students' lives. The fact that girls are reluctant to choose STEM subjects may be driven by the stereotypes that ability and interest in STEM are signals of masculinity. This is internalised by children and adolescents and reflected in their education choices. Explanations for ethnic differences in subject choice typically focus on cultural identity, stereotyping and discrimination. For example, there may be cultural differences in which subjects are considered more valuable (Archer & Francis, 2007), or teachers might have preconceived ideas about students' orientations to science based on their gender and ethnicity (Campbell, 2015). Moreover, the under-representation of women and people from ethnic minority backgrounds in science textbooks could have lasting negative impacts (Frost et al., 2005).

Amid these explanations, there are several concepts that could be operationalised quantitatively to shed light on intersectional differences in subject choices. For example, the concept of 'science capital' has been developed to understand students' engagement in science, defined as the extent to which their families have connections with or knowledge about science (Archer et al., 2012). White students and those from working class backgrounds have the lowest levels of science capital. The more prominent gender disparities in STEM choice among disadvantaged students may be a consequence of multiple barriers to science capital. While a working-class boy may grow up in a family with low science capital, they would also see themselves represented in science in the media, in textbooks, and be exposed to stereotypes about boys' relative competence in science. The negative impact of low science capital and stereotypes around class, academic capability and science suitability may thus be cancelled out. Working class girls, in contrast, would have no 'positive' stereotypes with which to override other barriers.

It could also be that class is directly related to the experience of gender, and to ideas about subjects that are suitable for boys and girls. There is some evidence that more educated mothers are more likely to hold egalitarian gender role attitudes (Farré & Vella, 2013), which may influence their children's subject choices (van de Werfhorst, 2017). Annette Lareau's (2003) seminal research highlighted the differences in parenting practices between advantaged and disadvantaged parents. Beyond relative differences in science capital, parents with more resources may be more able to combat stereotyping and foster their children's individual interests. Quantitative research exploring parents' gender role attitudes and parenting practices from an intersectional perspective could illuminate whether these factors play a role in the intersectional patterns of subject choice identified.

3.5.4 Educational and career aspirations

One possible explanation for inequalities in attainment and subject choice is students' aspirations, preferences, motivation, personality, and so-called 'non-cognitive skills' (Gutman & Schoon, 2013). Indeed, raising aspirations and improving pupils' confidence, motivations, and resilience are popular policy recommendations for tackling low educational attainment among disadvantaged groups (Sharples et al., 2011). Studies focusing on these traits are sometimes based on samples lacking ethnic and social diversity, a clear barrier to an intersectional approach. However, it is becoming more common for researchers to study concepts such as educational aspirations using nationally representative data (Goodman et al., 2011). Applying an intersectional framework to the analysis of aspirations and associated traits could shed more light on the intersectional patterns of attainment and subject choice described above.

Berrington et al. (2016) explored differences in students' aspirations to attend university as a potential explanation for attainment inequalities. Although their research did not identify any intersectional patterns in aspirations, it highlights the utility of studying intersectionality in relation to mechanisms that are thought to be key for educational attainment, alongside attainment itself. Moreover, the interaction between characteristics in relation to aspirations may be highly contextually specific, likely depending on differences in historical context. In contrast to Berrington et al., Howard et al. (2011) found interactions between US students' ethnicity and both social background and gender in determining career aspirations. For Native American and Asian/Pacific islander students, family income was associated with aspirations to enter prestigious careers, whereas for other groups this was not the case.

It is possible that differences in aspirations arise from students' realistic assessment of the barriers they will face when they leave schooling. It is well established in the literature that women and people from ethnic minority backgrounds are disadvantaged in the labour market, even when accounting for academic attainment (e.g. Crawford and Greaves, 2015). In England, women, people from lower income families and people from ethnic minority groups earn less upon graduation regardless of subject studied at university (Britton et al. 2016; Belfield et al. 2018). Students (and parents) may be aware of the additional barriers they may face and feel that they need to work harder and accomplish higher grades if they want to achieve a comparable position to more advantaged peers. Students who initially come from a more advantaged position in terms of labour market outcomes (for example, white, middle class boys) may be aware they do not need to work as hard. However, students' awareness of broader labour market inequalities is difficult to capture with quantitative data, and to our knowledge has not been attempted in large-scale, nationally representative studies. It should also be acknowledged that broader labour market inequalities and discrimination not only inform aspirations; they may also serve as a barrier to aspirations being achieved. Intersectional studies of educational aspirations should consider the role of both structure and agency in shaping how educational and career aspirations are formed and realised (Schoon and Lyons-Amos, 2017).

Furthermore, Strand (2014a, 2014b) suggests that some ethnic minority groups have greater resilience to lower socio-economic status because they possess 'ethnic capital'. Ethnic capital is a term coined to explain how attitudes towards education and a stronger work ethic within ethnic minority families leads to higher aspirations and attainment, especially when economic capital is low (Khattab, 2015; Modood, 2003; Strand, 2014a). This may operate through several mechanisms, for example selective immigration of highly motivated individuals, or as a response to the labour market discrimination discussed above. This is particularly important considering the differences in associations between social background and performance *across* various different ethnic minority groups.

Ethnic capital requires further investigation in quantitative research, perhaps by measuring social background along different dimensions, including education level of parents or social position before immigration. Research could also explore the impact of other factors associated with ethnicity, such as generation of immigration (e.g. Lessard-Phillips & Li, 2017). Interestingly, patterns in the US are very different. Alon (2007) shows that the effects of disadvantage are far worse for black students than for white students. Researchers could exploit these cross-national differences to help pinpoint mechanisms. For example, differences in the impacts of social background by ethnicity may in part be explained by different policy responses to multiculturalism, or differences in immigration patterns and forms of discrimination. Also, more work needs to be done to analyse different patterns of 'non-cognitive skills' and resilience across multiple ethnic groups, rather than a binary comparison of white versus non-white.

3.5.5 The importance of context

Most of the studies reviewed have focused on either the US or UK, and few quantitative studies have addressed the contextual specificity of intersectional inequalities. However, situating intersectional inequalities in their institutional context could help to explain how and why they occur. Part of the definition of intersectionality is that inequalities are contextually specific (Browne & Misra, 2003; Crenshaw 1989; 1981; Gross et al., 2016b). The characteristics and practices of schools and universities, such as programme structure and the tracking of students into different educational pathways based on their abilities or interests, shape young people's routes through the education system (Charles and Bradley, 2004; Frenzel et al., 2010; Kutnick et al., 2005; Mann & DiPrete, 2016). With multi-level data including school or university information, researchers could explore whether these institutional practices are also associated with intersectional gender, ethnic and SES differences in attainment and subject choice.

This chapter has noted some key differences between countries, which could be explored further. In terms of educational attainment, the key disadvantaged groups in the UK are socio-economically disadvantaged white and black Caribbean boys, whereas in the US, black male students are particularly disadvantaged. These cross-country differences could be related to several factors, including history, culture, politics, or institutions. Future research exploring cross-national differences in intersectional inequalities could build upon existing research, which has identified, for example, that more standardised education systems promote social background and ethnic equality (Montt, 2011; van de Werfhorst & Mijs, 2010; Pfeffer, 2008), and that male over-representation in STEM fields of study in higher education, and gender differences in aspirations for STEM study, are particularly pronounced in more economically advantaged nations (Charles & Bradley, 2009; Charles, 2017). On a smaller scale, regions within countries could be compared.

These considerations suggest that cross-country or regional patterns of intersectional gender, ethnic and SES differences in attainment and subject choice would be a fruitful area for future research. An intersectional approach therefore has great potential to illuminate the links between social structure and a combination of individual characteristics in determining educational outcomes (Gross et al., 2016b). Given the availability of representative longitudinal cohort studies, the charting of intersectional inequalities over the educational life-course and across cohorts is another clear next step and will be a vital addition to our understanding of when and how intersectional inequalities emerge, as well as how they are changing across successive generations.

3.5.6 Descriptions of intersectionality

As noted previously, few of the studies we have outlined explicitly refer to intersectionality as a theory, method or hypothesis. While some studies do mention intersectionality theory (Strand, 2014a, Berrington et al., 2016, Codiroli Mcmaster 2017), many simply note the reasons there may be an interaction along a particular axis of inequality. This raises the question of whether the studies described can be considered fully ‘intersectional’. Moreover, some may take issue with studies referring to intersectionality without empirically considering structural factors and systems of power that give rise to inequalities (Gillborn et al., 2017). While recognising these critiques, we believe that the studies discussed still constitute an important step in our understanding of intersectional inequalities, and should not be dismissed simply for not applying the theory comprehensively. Not only do these studies improve the description of educational inequality, they also identify many areas for further investigation.

Gross et al. (2016b) suggest that the need for empirically verifiable hypotheses in most quantitative research hampers the explicit application of intersectionality. For such hypotheses to be developed, relevant interactions need to be specified in advance and justified theoretically. Although this approach is less common and more challenging, we wish to draw attention to quantitative studies that have made progress in this direction by providing a more explicitly intersectional framing of their analyses and results. A recent study by Van de Werfhorst (2017), on gender differences in fields of study, sets out to test an intersectionality hypothesis, supported by an in-depth discussion of why the influence of gender may vary by social background. He also considers contextual factors influencing this intersectional hypothesis, by exploring changes over time. He finds that, over the period 1931-1989, gender segregation into fields of study decreased, and the relationships between gender, social background and field of study also changed over time. Being more explicit about the use of an intersectional approach not only makes the research easier for other academics to discover and synthesise, but also facilitates better interpretation of results alongside theoretical work. We believe that more quantitative researchers should be taking this type of explicit approach. However, studies can be even more overt than this, by incorporating broader knowledge about where specific intersections are likely to be found as part of the formulation of hypotheses, rather than in a post-hoc discussion. In this way, studies can go beyond superficial use of the intersectionality concept.

3.6 Challenges and innovations

While some scholars have argued that the rigid nature of quantitative research masks the truly complex relationships between individuals' characteristics and outcomes (Trahan, 2011), we have described an emerging body of quantitative educational research that operationalises intersectionality in compelling and impressive ways. However, there are some methodological difficulties with applying an 'inter-categorical' approach to quantitative research on educational inequalities. The first concerns the categorisation of individuals into pre-defined groups. This could obscure the true relationship between individuals and power structures within society and will undoubtedly lead to mis-classification of some individuals, who may face more or less disadvantage than the findings suggest. For example, a person is not just female and from a working- class background, but many other things besides. Indeed, a fundamental aspect of the intersectional approach is to question the very nature of categories such as gender, ethnicity, and class (McCall, 2005; Gross et al., 2016b).

Recent methodological innovations in survey research can mitigate the categorisation problem to a certain extent. For example, Next Steps contains detailed indicators of parents' and neighbourhood characteristics that can be combined to construct a multi-dimensional measure of social background (e.g. Anders, 2017; Codiroli McMaster, 2017). These include parents' occupation, education, entitlement to Free School Meals (FSM), home ownership and neighbourhood deprivation. Next Steps also contains measures of aspirations and attitudes, which can be explored as potential explanations of inequalities. Earlier cohort studies can also be used to analyse the multi-dimensionality of social background (parental class, status and education) and its effect on educational outcomes (Bukodi and Goldthorpe, 2013). Furthermore, UK longitudinal studies often over-sample ethnic minority groups, as does the German National Educational Panel Study, meaning that robust conclusions can be drawn, as there are sufficient numbers of cases available. Finally, the move to increasingly link survey data with administrative sources, such as tax and health records, will be hugely beneficial for research into intersectional inequalities.

But despite the rich data available for studying intersectional inequalities in education, further innovation is needed. Most large-scale surveys do not oversample on characteristics that are relatively uncommon, but which impact educational outcomes. Only recently have longitudinal studies over-sampled people from ethnic minority backgrounds, and many older birth cohorts (for example the British Cohort Study (BCS70), initiated in 1970, and the National Child Development Study (NCDS), initiated in 1958, do not have large enough samples to allow complex analysis of differences by ethnic group.

Important aspects of inequality can be overlooked because of data limitations. For example, despite policy interest in students with caring responsibilities and the influence of these responsibilities on educational trajectories (Department for Education, 2016), this has not, to our knowledge, been explored in large-scale quantitative research. Nor have we been able to find any quantitative research that considers the intersectional experience of students whose gender identity differs from that which they were assigned at birth, or parents and children with disabilities. The information is often simply not collected, and where it is, sample sizes are too small. Studies considering the experiences of smaller (yet very significant) groups of students would benefit from more targeted data collection, and researchers can do more to inform the data collection process by suggesting that the necessary questions are asked when survey questionnaires are in development.

A second potential problem concerns the statistical methods used to identify intersectional inequalities in quantitative analysis, which were described earlier. The use of interaction effects is not always straightforward in non-linear regression models, which estimate the probability of an outcome or event occurring, such as logit and probit models. As Ai and Norton (2003) point out, the coefficient on an interaction term is not easily interpreted in such models, and the true relationship could even go in the opposite direction (positive or negative). Researchers therefore need to be careful about how they present results. For example, instead of just reporting coefficients, researchers can construct charts to visualise the marginal effects of relationships between the focal variable and outcome, broken down by the moderator variable, and assess the direction and extent of any relationships. Moreover, there are limitations on the number of interaction terms that can be included in quantitative research from a practical point of view. For example, the inclusion of 10 dimensions of inequality would lead to 1013 possible interaction terms. Researchers therefore need to be careful about the categories they choose to focus on and the way they present results.

Another way for researchers to avoid assigning individuals to predetermined groups, and to avoid the pitfalls associated with interaction terms and sub-group analysis, is by using latent variable methods. Latent variables are hypothetical constructs that are measured quantitatively using multiple manifest indicators (Bollen, 2002). For example, social background could be operationalised using a combination of parents' education, parents' income and access to cultural resources in the home or community. One could then see whether gender or different categories of ethnicity are statistically associated with a particular combination of disadvantages. Latent variable methods could also be used to explore complexity within a given social category (for example, pupils on free school meals), operationalising what McCall (2005) terms an 'intra-categorical' approach to intersectionality.

Although latent variable methods are not always informed by an intersectional approach, the methods are well suited because they emphasise the complexity and configurations of characteristics[3]. They also do not impose assumptions, instead allowing patterns to emerge from the data. An example of this is a study by Alon (2007), which uses latent variable techniques to analyse inequalities in college graduation. Alon finds that multiple social, economic and academic disadvantages interact in complex configurations, and have a combined effect on students' graduation likelihood, which is also moderated by gender and ethnicity. While one needs to be careful about the extent to which complex combinations of characteristics identify meaningful groups, latent variable methods are a promising and currently under-used quantitative method for studying intersectional inequalities in education.

Presentation and framing of analysis is key in communicating results from quantitative studies focusing on interactions between characteristics, particularly when relaying results to audiences less experienced in interpreting quantitative research. Academics should always be mindful of which groups they are foregrounding, which groups are being sidelined, and the political and policy

[3] It should be noted that this approach still requires categorizing individuals as a first step, so would still not be fully intra-categorical in the way described by Gross et al. (2016b) and McCall (2005).

implications of those decisions. For example, the foregrounding of white working class boys in some studies has drawn policy attention to this group at the expense of other groups. Another example of this, not in the field of education, is a recent highly publicised study by Chetty at al. (2018), which focused on the outcomes of black men compared to white men from similar social origins, arguing that women were not affected to the same extent by racial inequalities. However, this conclusion rested upon the particular comparison they were making (black men versus white men) and the outcome they chose to focus on (income). Researchers should be careful to be explicit about what can and cannot be inferred from their research, based on the methodological decisions they have made.

While we are optimistic about the application of intersectionality within quantitative studies of educational inequality, we do recognise the limits to this approach. As Gross et al. (2016b) argue, quantitative research is less well placed to investigate the 'anti-categorical complexity' aspect of intersectionality. Interrogating the nature of social categories requires recording individuals' subjective experiences and capturing concepts such as discrimination, stereotyping and prejudice. These concepts can be challenging to measure using quantitative data. For example, nuanced measures of the experience of discrimination are rarely available in survey datasets (Harnois, 2013), and it is difficult to capture subjective identity in large-scale, quantitative data. Anti-categorical complexity is therefore best suited to a qualitative research approach and there are many good examples of this, such as Stahl's recent work on subjective ideas of masculinity, class belonging and education among working-class boys (e.g. Stahl, 2017).

3.7 Discussion

Educational inequalities are a major challenge for policy makers, educators, students and their families. In this chapter, we have described the current status and main contributions of quantitative intersectional research on inequalities in educational attainment and subject choice. We have highlighted important findings from this literature, discussed why the approach is important and considered future innovations that would help strengthen the contribution of intersectionality to quantitative research on educational inequality.

While intersectionality theory is more commonly associated with qualitative research, quantitative researchers are increasingly applying it to their research into inequalities. The increasing availability of large-scale survey and administrative data has facilitated the study of more complex social identities, and we have outlined a number of statistical methods researchers have employed in analysing such data. The majority of these studies take an 'inter-categorical' perspective on intersectionality, focusing on the interactions between gender, social background and ethnicity, and their combined influence on outcomes. Some also take a broader intersectional perspective on inequality, emphasising multi-dimensionality and contextuality.

The research reviewed in this chapter shows that gender, social background and ethnicity influence educational outcomes in complex, intersecting ways. Researchers should be mindful of these intersections when conducting research into the themes of educational attainment and subject choice. Specific intersections have been highlighted as particularly important. Firstly, socio-economic disadvantage has different effects on educational attainment and subject choices depending on gender and ethnicity. For ethnicity, although inequalities can sometimes be 'explained' by the unequal distribution of socio-economic resources across ethnic groups, this is not always the case. In the UK, some ethnic minority students seem more resilient to the effects of disadvantage. Patterns emerging from the combination of ethnicity and social background are different across countries.

Gender differences also seem to be intertwined with social background: working class boys have the lowest attainment, and less advantaged female students are least likely to study STEM subjects in higher education in the UK (but *most* likely to in the US). We noted that these findings are primarily descriptive, and that by focusing on psychological drivers of attainment, considering comparative and historical context and incorporating further categories representing different types of disadvantage, quantitative intersectional research into educational inequalities can make a stronger contribution. Some progress has been made in this direction, but further work is needed. Also, it is likely that gender, social background and ethnicity interact in predicting additional outcomes that have not been covered in this chapter, but may be equally important; for example, early years development (Walker et al., 2011), and university completion (Crawford, 2014).

The chapter highlighted several challenges associated with applying an intersectional approach to the quantitative study of educational inequalities. We suggested that these challenges are not insurmountable but require a creative approach and more data resources. For example, although the problem of allocating individuals to pre-defined groups cannot be fully resolved, using multi-dimensional measures of social background and other characteristics can mitigate it. We also suggest that researchers should be careful about the presentation and interpretation of results, and look into techniques such as latent variable methods to analyse the complexity of inequalities. While an 'anti-categorical' approach may be most suited to qualitative research, there is a clear gap in the quantitative literature concerning an 'intra-categorical' approach to intersectionality, analysing disparities *within* social groups.

We have several recommendations for the future of intersectional, quantitative research on educational inequalities. Firstly, researchers who are interested in studying these complex inequalities should explicitly engage with intersectionality theory, making sure that the intersections they choose to target are well grounded in theory and prior research. It is challenging, but not impossible to develop empirically verifiable hypotheses concerning intersectional inequalities. However, it requires engagement with theory and empirical findings beyond one's immediate disciplinary and methodological bubble. We believe that, if quantitative researchers do this, they can tap into the unrealised potential for intersectionality in quantitative research. Moreover, their research can have a deeper impact, not least by helping to facilitate more inter-disciplinary, multi-method dialogue in educational research.

Secondly, to facilitate a more thorough application of intersectionality to the quantitative study of educational inequalities, the survey and administrative data that is the basis of much quantitative research in education must include more detailed aspects of social location and identity. This will require close working relationships between academics, civil servants, policy makers and data controllers to ensure rich data is available for analysis without jeopardising the privacy of participants. This requires all parties' acknowledgement that intersectional research can make a meaningful contribution to tackling educational inequalities. In the UK, some steps have been made to facilitate this by increased access to linked administrative datasets, which will also help with the analysis of smaller demographic groups. However, there remains a long way to go (Economic and Social Research Council, 2017).

Thirdly, we suggest that more attention should be paid to comparative and longitudinal aspects of intersectional inequalities in education. Quantitative researchers need to go beyond identifying intersectional inequalities, by distinguishing the specific historical and policy context in which they arise. There are several potential challenges here. Practically, the quality of data available in survey and administrative datasets varies across countries, and identifying whether differences in associations arise from genuine intersectional inequalities, or to measurement differences, will be challenging. Furthermore, it will be difficult to pinpoint the reasons for differences in intersectional inequalities across contexts and over time. Nonetheless, this work could help to inform policy and practice aimed at ameliorating these damaging educational differences, along with enhancing our understanding of systems of power and how they have evolved over time to privilege and disadvantage particular groups.

The value of the research described in this chapter is, first and foremost, to improve the description of inequalities, showing that 'educational inequality' is not one phenomenon, but many. Although not all the studies discussed explicitly engage with intersectionality theory, they still make a valuable contribution to the field of research on intersectionality and educational inequalities and identify many areas for future research. The approach can also offer explanations of intersectional inequalities and ways to address them. Quantitative researchers now need to go further by embracing intersectionality theory, along with the insights of qualitative research, and using it to develop and test explicitly intersectional hypotheses. While it is still imperative to recognise the overriding impact of singly-defined characteristics such as ethnicity (Gillborn et al. 2017), we trust that this chapter will motivate quantitative educational researchers to apply the concept of intersectionality in their work. We hope that it will become common practice (where there is justification to do so) to test for interactions when considering inequalities both within and across countries, and over time, motivating and contextualising this approach using intersectionality theory.

Chapter 4: Data and samples

4.1 The benefits of using representative, longitudinal data

Analysis conducted in this thesis used data from three main sources. The first two empirical chapters use data from Next Steps (formerly the Longitudinal Study of Young People in England (LSYPE)), and the third empirical chapter uses data from the British Cohort Study (BCS70) and US data from the National Longitudinal Study of Youth (NLSY79).

These datasets are all longitudinal surveys, with the same individuals interviewed across a number of years. This has allowed me to test associations between child, adolescent and adult characteristics. Relying on participant recall to test associations would have caused numerous problems with interpretation of results. For example, if students were asked about their feelings towards subjects after they had made their university choices and received their final exam grades, their responses would likely be influenced by these outcomes. A young person may choose a subject based on outside factors (for example parental influence), and either gain an appreciation of the subject later or incorrectly recollect a prior personal preference. By being able to see responses of young people with no prior knowledge of the path their lives would take, I can circumvent some of these issues with reverse causality. Of course, this is not perfect, as its likely children would have received feedback in earlier childhood about their performance which would be correlated with later ability and choices, however their specific choices would not have influenced perception of previous interest or ability.

The studies all contain a wide range of information including education, economic activity, health, childhood and background. The nature of these datasets allowed exploration of more complex associations between background characteristics and both educational choices and occupational outcomes. The rich information included in these surveys was invaluable for testing associations between a broad range of social background characteristics, and in understanding the associations between psychological characteristics and later destinations. Whilst administrative data available could help to answer some descriptive aspects of the questions posed, for example the relationship between gender or ethnicity and subject choices, it could not be used to study the more complex questions.

Nonetheless, where appropriate and possible, data from the Higher Education Statistics Agency (HESA), linked with the National Pupil Database (NPD), is used to test the representativeness of findings presented in this thesis. This analysis is presented later in this chapter. HESA data includes data from all publically funded higher education institutions across the UK on individual students course of study and university, and basic demographic information. The NPD includes exam and attainment records of all students in state-funded education in England. Data are held by HESA and the Department for Education (DfE) and are released only to authorised researchers under approved conditions.

Another benefit of the longitudinal surveys used is that they are all representative of the population being studied; at least to the extent that this can be measured through observable variables. Where non-response, early dropout, or the need to oversample certain under-represented groups compromises representativeness, survey weights account for the effects this may have on the study populations. This allows me to make stronger inferences about the applicability of findings to the population as a whole, and reduces the possibility that findings are unique to study participants. As noted, there remains a prospect that the study populations differ from the general population in some *unobserved* characteristics, and this would affect the generalisability of results if these differences were also associated with the outcomes of interest and independent variables in statistical models. Given the absence of more detailed administrative data, however, the studies used give the best approximation of the general populations characteristics.

4.2 The Centre for Longitudinal Studies

Two of the cohort studies used in this thesis, Next Steps and the BCS70, are held by the Centre for Longitudinal Studies (CLS), a research centre based at UCL Institute of Education and funded by the Economic and Social Research Council (ESRC). The centre runs and manages the studies, including the coordination of data collection, testing of new methods and techniques to increase response rates, cleaning and uploading data, and generating many derived variables used by researchers. They also provide detailed user guides to allow researchers to navigate the datasets effectively.

4.3 Next Steps (LSYPE)

Next Steps is a national longitudinal study that followed the lives of around 16,000 young people. The study began in 2004 and remains ongoing. Initially funded and run by the Department for Education up to 2012, the study moved to CLS and has been funded by the ESRC thereafter. When the study began, participants were between 13 and 14, and the most recent survey was conduced when participants were between 25 and 26. This thesis draws on responses to interviews in the first wave, and in wave seven, when participants were between 19 and 20. The study was linked with the NPD, giving detailed information on participants' attainment at school.

Table 4.1: Summary of Next Steps response rates (Waves 1-8)[4]

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Year	2004	2005	2006	2007	2008	2009	2010	2015
School year	9	10	11	12	13			
Age	13-14	14-15	15-16	16-17	17-18	18-19	19-20	25-26
Key stage	KS3		KS4 (GCSE)		KS5 (A level)			
Sample	15,770	13,539	12,439	11,801	10,430	9,799	8,682	7,707
Response rates	74%	86%	92%	92%	89%	87%	90%	51%
Mode	Face to face interview				Mixed (depending on preference): face to face, online or over the phone			

4.3.1 Is the Next Steps university sample representative of the English university population in 2010?

Previous research has shown that Next Steps data is representative of the population of the time, however participants in the sample were more likely to attend higher education by wave seven. Less is known about the representativeness of the sample of students who *did* attend university and higher

[4] Adapted from: http://www.cls.ioe.ac.uk/page.aspx?&sitesectionid=1248&sitesectiontitle=About+the+sample

education to study for a degree. Table 4.2 outlines the characteristics of the whole Next Steps sample in 2010, the HE and university samples, and HESA data including all UK first year students. Proportions are weighted to account for longitudinal and cross sectional non-response, and over-sampling of ethnic minorities.

We would expect differences in characteristics of the whole Next Steps sample and the HE/ University sample, particularly that there will be higher proportions of women and students from more advantaged families engaging with tertiary education. We may also expect young people from ethnic minority backgrounds to be more likely to attend university than white young people (Crawford & Greaves, 2015). This is generally reflected in the data. The gender split for the whole sample is close to 50/50, however 55% of the Next Steps HE/ university sample are women. HESA estimates that 57.5% of first year university students were women, suggesting a slight over-representation of men (or under-representation of women) attending university in the Next Steps sample. This may be driven by the fact that overall students within the sample were more likely to attend university, and young men who were not likely to aspire to university may have been least likely to respond to the survey at all.

For ethnicity we also see higher proportions of most ethnic minority groups (particularly Indian and black African individuals) attending HE, and lower proportions of white individuals. These proportions generally reflect HESA estimates, however there's some indication that the Next Steps has underestimated the proportions of black young people, and overestimated proportions of Indian and Bangladeshi young people attending university.

HESA has not published information on students' social background and attendance at university, but in line with expectations the HE sample have more educated parents than the overall sample. Just fewer than 18% of Next Steps participants had at least one parent educated to degree level, compared with 33% of the HE sample.

Table 4.2: Characteristics of students in the main Next Steps samples compared with HESA data.

Characteristic	Whole Next Steps wave 7 sample	HE (University) final sample	HESA estimates[5]
		%	
Sex			
Female	49.4	55.4 (55.4)	57.5
Male	50.6	44.6 (44.6)	42.5
Ethnicity			
White	86.6	81.6 (81.3)	80.0
Mixed	2.6	2.8 (2.8)	3.0
Indian	2.5	4.6 (4.7)	3.4
Pakistani	2.3	2.8 (2.8)	2.3
Bangladeshi	1.1	1.4 (1.5)	0.8
Black Caribbean	1.1	1.1 (1.1)	1.8
Black African	1.6	2.5 (2.5)	4.7
Other	2.2	3.3 (3.3)	4.0
Parents education			
Degree or higher	17.7	32.7 (33.4)	N/A
HE or A-levels	34.2	36.2 (35.8)	N/A
GCSEs or lower	48.1	31.1 (30.9)	N/A

4.4 BCS70

The British Cohort Study is another longitudinal panel study. In contrast to the Next Steps survey, participants were followed from birth rather than adolescence. All babies born in a single week in 1970 were eligible for inclusion in the survey, and data were collected from parents, midwifes and teachers as well as the cohort members themselves. The study is ongoing; with fieldwork for the 2016 survey completed late 2017. Chapter seven draws on data from all waves, including family background, childhood information and cognitive scores in early waves and income information in later waves. Many variables used were not included in analytical models or descriptive statistics, but were used in the construction of weights or in the multiple imputation models.

[5] Estimates include first year students in England

Table 4.3: Summary of BCS70 response rates (Waves 1-8)[6]

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9
Year	1970	1975	1980	1986	1996	2000	2004	2008	2012
Age	0	5	10	16	26	30	34	38	42
Sample	16,571	13,071	14,874	11,621	9,003	11,261	9,665	8,874	9,842
Response rates	95.9%	79.0%	88.9%	70.6%	55.9 %	71.5%	75%	75.6 %	74.6 %
Mode	Face to face				Postal	Face to face		Phone	Face to face

4.5 NLSY79

The NLS79 differs from the former two studies in that participants surveyed are of varying ages. Individuals born between 1957-64 were followed, and the survey started in 1979 when individuals were between 14 and 22 years old. They were followed annually initially, giving rich information over relatively short intervals, and biannually after 1994. Information was gathered covering a number of topics including education, employment, household and contextual variables, family structure, income, and non-cognitive and cognitive traits. The study began with 12,686 respondents and has achieved high response rates each year, with 9,964 remaining in the study in the 26th iteration (in 2014). This thesis draws on responses from individuals up to 2012, however the main analysis is conducted with a sample that remained in the survey up to 2004.

4.6 Measuring family background

In attempting to understand the relationships between social background, education, and occupational outcomes, researchers have used a number of different approaches in measuring social background. These primarily include income, social class, and parent's education. Parental income generally measures family's financial standing, giving an indication of the additional educational and time resources they have to help their child succeed. This includes paying for tutors and extra curricula activities in childhood, supporting offspring in adulthood to pursue goals, and

[6] Information available from: http://www.cls.ioe.ac.uk/

protection from the negative impacts of financial stress. This is a reasonable proxy for overall resources, however studies rarely take account of debt, expenditure and additional family costs, or whether parents do choose to invest money in their children.

In contrast, social capital approaches aim to take account of the stability of ones position, and a collection of attitudes and behaviours that may help children succeed. These are primarily measured through parents' education level and occupational social class, where people who attain higher levels tend to 'move in different circles' to people at lower levels. In this thesis, NS-SEC classifications are used to measure occupational class (DiPrete, Erikson, & Goldthorpe, 1993).

Although they are both strongly associated with family income, greater educational ambitions, interest and engagement with pursuits that compliment school learning can help children succeed even without a large disposable income. Parents' education specifically may signal better knowledge of educational systems and how to navigate these to their children's advantage.

The extensive amount of variables collected in all three studies used in this chapter, in particular in Next Steps, has allowed me to explore which measures are most appropriate for analysis in this thesis. In chapter five I set out to directly test which measures were most associated with field of study, finding that, when accounting for social class and education, income had little association with choices. Overall, it appeared that parent's education level was most predictive. I also pursued other approaches to measuring background, for example deriving a composite score that included all three measures, and additionally parents' perception of their financial position (in an attempt to include consideration of debt or additional costs which may mean income does not reflect true economic position).

Consideration also needed to be given to cultural differences in conceptualisation of social position. Occupational class is not a commonly used measure of status in US literature, where research tends to focus on income or education level. Class as a 'culture,' or set of attitudes and values is almost by definition opposed to the American Dream narratives of individualism, where anyone can succeed based on merit alone. A persons background remains strongly related to their future chances in the UK and the US (Jäntti et al., 2006), but the system of measurement of that early position typically differs. Despite this, I have attempted to code parents' occupation into levels of prestige, using Duncan's Socio-Economic Index (SEI). The UK NS-SEC classifications were developed incorporating job stability and security, however the SEI only takes account of average income and education level. The decision was made not to use this measure in final analysis because of differences with the UK measures, and because education level was more comparable, however they are incorporated into imputation models.

4.7 Measuring subject studied in NLSY79

In Next Steps and BCS70 participants who were attending university or had stated that they had obtained a degree were asked their subject of study in the most recent wave of data used in this thesis. In NLSY79, however, participants were asked in each wave of data collection which subject they were studying (if they were attending university), and whether they had obtained a degree. Participants did not respond to the question if they had not yet chosen a major, however they could have chosen a major and later switched to a different major (particularly between the first and second year of study). To construct a variable for field of study, I created a variable that included the most recent (or only) response to the questions about field of study and degree attainment (including 4 year or 2 year degree). I then restricted the sample to those who reported studying for a 4-year degree and obtained their degree.

4.8 Missing data strategy

Survey datasets, particularly longitudinal survey datasets, usually suffer from the problem of missing data. Data on variables may be missing because participants did not respond to particular items in questionnaires, for example if the question was sensitive or poorly worded, and they chose not to give an answer. Data may also be missing because a participant left the survey altogether, and did not wish to participate in later surveys (or could not be traced for follow-up). The former type of missing data may lead to biased estimates if the characteristics of those who did answer the question differed from those who did not, and if these characteristics were also related to the outcomes in question. The latter form of missing data, or attrition, may compromise the representativeness of the survey. Whilst each of the surveys aimed initially to create a sample that was similar to the overall population of each country, it is likely that particular groups would be more likely to attrit over time. People who become homeless, for example, would be harder to contact over multiple waves, and thus some of the most disadvantaged individuals would no longer be represented. There may also be missing or incorrect responses through human error, if participants/ interviewers skipped over a question by mistake or input an incorrect response.

There are three main types of missing data; Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not at Random (MNAR). Data MCAR would cause the fewest problems to analysis, as the missingness would be uncorrelated with any variables of interest in the study. This may occur if the missingness was driven purely by chance. If data were MCAR, analysis using listwise or pairwise deletion methods (or complete case analysis) would be justified, as the only implications for results would be a loss in sample size and thus power. If this was not the case, however, using these approaches would bias results and it would need to be stated that the sample is no longer representative. Whilst it may be the case that missing data through error would be random, it would be too strong an assumption to assert that all instances of missing data were completely random.

In this thesis I have used missing data strategies that assume data are MAR; that is, that missing responses or attrition can be fully explained by observable variables included in the dataset. I therefore also assume that there are no unobserved variables that explain some of the missingness, for example underlying motivation, and that the variables included in the model measure each construct reliably. This assumption becomes less problematic because of the rich set of indicators included in each dataset, yet is more conservative than assuming data are MCAR. Data may also be MNAR, for example, if the missingness is directly related to the construct being measured. This may occur, for example, if father's education level is missing because the participant does not have a father. Where possible, this is taken account of when coding the data, and an additional level created. For other variables this is not possible, for example if people who do not like science are more likely to skip questions about their attitudes to science.

Given I assume data are MAR, I use two main missing data strategies throughout this thesis. The first, to account for cross-sectional and longitudinal non-response or attrition, is weighting. This strategy assigns greater weight in analysis to individuals if their characteristics are associated with attrition. Individual responses are given less weight if people with their characteristics are *more* likely to either remain in the study or to be sampled in the first place. In Next Steps, certain ethnic minority groups are over-sampled, to allow researchers greater power in analysis including these groups. In this case, weighting assigns less weight to these individuals. In Next Steps and the NLSY79 weights are provided for researchers, and these are used in all analysis. In BCS70, however, weights were not provided. I therefore constructed weights using logistic regression methods, predicting probability of being in the most recent wave (2012) based on baseline characteristics. Characteristics chosen were informed by Mostafa & Wiggins (2014), and included sex, birth weight, parity, mother's age, whether mother lived in the southeast of England in the first survey, social class at birth, and mother's and father's age at completion of education.

Whilst this helps to account for attrition over time and non-response to the entire survey, it does not account for missing data on particular items within the survey. To account for this, I constructed multiple imputation models through chained equations for each of the studies. This method is considered particularly preferable when data are MAR (Allison, 2001), and to lead to less bias in results than complete case analysis, or other imputation methods. Simple imputation methods underestimate standard errors, overestimate 't' statistics and can therefore return significant effects where there aren't any. This is because by imputing a single value for a large proportion of cases (for example the mean) the variance in scores appears artificially smaller. If there were responses on these cases, even if data were MCAR, there would likely be random variation between respondents. Single regression imputation methods would also return biased standard error and test statistic estimates. When using regression methods the imputed data become a direct function of the outcome and other predictors in the model. This then artificially inflates the relationship between the imputed variable and the outcome. MI deliberately introduces this random variation by creating many datasets based on the regression equation entered, and takes the mean value from these estimates. Variation across the different imputation datasets is then utilised to calculate standard errors and test statistics that are larger and smaller respectively to reflect expected natural random variation in responses. The standard errors are thus a function of the variance between cases within each data set, and the variance between datasets. In all MI models, I created 20 datasets, as guidance suggests using a large number of imputations (Graham, Olchowski, & Gilreath, 2007). Analysis was conducted using the MI ICE command in STATA (Royston, 2004). Variables included in each imputation model are shown in appendix A.

Chapter 5: Who studies STEM subjects at A level and degree in England? An investigation into the intersections between students' family background, gender and ethnicity in determining choice

5.1 Introduction

There is a long-standing skills gap in the supply of graduates with much-sought-after expertise in STEM (Science, Technology, Engineering and Mathematics) subjects, causing concern for how economies will cope with our increasing dependence on technology in everyday life (Winterbotham, 2014). A rich literature has emerged, with policy-makers, academics and stakeholders in industry working to further understand the full extent of the problem. The Social Market Foundation has identified an existing shortage of up to 40,000 workers with STEM skills, and considering trends in industry it is predicted that this will increase significantly if steps are not taken to close the gap (Broughton, 2013). A particular problem is that socio-economic background, gender and ethnicity are all associated with the study of STEM subjects (CaSE, 2014; Equality Challenge Unit, 2014).

The economic case for increased participation and diversity in STEM fields is clear, but there are also substantial benefits to be had for individuals. For example, those who study STEM subjects at degree level and General Certificate of Education (GCE) Advanced Level (A level) typically earn higher salaries later in life (Dolton & Vignoles, 2002; Greenwood et al., 2011). Despite this, the problem of low uptake seems a particularly large concern in the UK, which has one of the lowest shares of 15-year-olds aspiring to pursue STEM careers of OECD countries (OECD, 2012). In the interests of the promotion of social mobility and equality of opportunity, it is important that individual benefits are not restricted by a student's social background, gender or ethnicity. Recent policy changes have led to an increase in post-compulsory mathematics qualifications available (Department for Education, 2014), which may contribute to increased basic skills in maths, however, they may not necessarily lead to an increase in participation at degree level. It is therefore important to understand which students do not study STEM subjects, and why particular groups have lower participation.

Prior research in the area has considered reasons for decreased participation in

STEM subjects for all students, often with particular focus on gender disparities. Reasons put forward for lack of engagement include students' values, perceptions of the importance and relevance of STEM, shortages of maths and science teachers, perceptions that STEM subjects are more difficult or 'boring' compared with other subjects (Wynarczyk & Hale, 2008), and teaching methods and styles (e.g. Gilbert, 2006; Pampaka et al., 2012a,b). In response to decreasing participation, a large research initiative—the Targeted Initiative on Science and Maths Education (TISME)—was set up in the UK. Key findings from five large-scale projects included that the perception of ability and knowledge of usefulness of STEM appeared to drive issues with uptake, rather than interest in or enjoyment of science (TISME, 2013). Furthermore, science capital in families was an important driver of choice; students whose parents were engaged with STEM or worked in STEM careers were more likely to study STEM further (Archer et al., 2012). There is less research, however, on how these mechanisms relate specifically to student characteristics, especially in respect to students' background and ethnicity. An important prerequisite to understanding exactly which mechanisms lead to decreased engagement amongst particular groups is to fully understand which student characteristics are associated with choice, and how.

5.1.1 Family background, gender, ethnicity and subject choice

Family background is a key predictor of students' academic progress; a strong association persists between income and achievement across subjects in the UK (see The Royal Society, 2008). In consideration of this relationship, there is a growing literature detailing how this translates into access inequalities in Higher Education (HE) (e.g. Gayle et al., 2003; Blanden & Gregg, 2004; Anders, 2012), however, the question of subject choices is relatively under-researched in the UK. The Royal Society identified prior attainment as the strongest predictor of subject choice (The Royal Society, 2008), and considering there are large differences in attainment by students' background, it is possible that disparities in uptake by social position reflect these academic disparities.

Research in the UK reveals some association between family background and subject choice. Van de Werfhorst et al. (2003), using the 1958 British Birth Cohort Study, found that social class was related to choice of prestigious fields of study, including medicine and law, at university. Focusing on STEM subjects directly, Gorard and See (2009) showed a clear disparity in numbers of students choosing to study STEM subject post-16 by Free School Meal (FSM) status (a measure of disadvantage based on students' family income; students whose parents earn below a certain threshold are eligible for free school lunches in the UK). Although lower attainment amongst students eligible for FSM was shown to be an important reason why they may be more reluctant to study STEM, the authors argue that this does not fully explain disparities by levels of advantage. Research into students' background and science participation has shown that students' social class is associated with science capital, which would lead us to expect students' background to be positively related to participation (Archer et al., 2012). It is clear, however, that the relationship between background and uptake, given prior attainment, has yet to be fully unpicked.

Sociological theory offers some insight into why educational inequalities by students' social background emerge. According to Boudon's (1974) model of relative risk aversion, extended formally by Breen and Goldthorpe (1997), individuals will aim for a social position that is at least as good as their parents', with the key motivation of avoiding downward mobility (Breen & Yaish, 2006). The theory's implications for vertical stratification are clear; students from higher socio-economic status (SES) backgrounds would be more likely to attend university, as this will be necessary for maintaining their social position. For horizontal stratification, however, the picture is less clear. On the one hand, students from higher SES groups may be more concerned about choosing subjects with higher returns upon graduation (including STEM subjects). For students from more working-class backgrounds, or with parents having few qualifications, by studying any subject at A level or university they will be moving up the social ladder. In accordance with this interpretation, Davies et al. (2013) found that students from higher socio-economic backgrounds were more concerned with financial returns when making educational choices. Conversely, the theory could suggest that more disadvantaged students will be more concerned with returns to subjects than their peers. For students from lower SES groups, there may be more risks associated with the study of arts and humanities subjects. More advantaged students will usually have more networks to draw on after graduation, and may be able to receive more financial help from parents when gaining additional work experience (for example, through unpaid internships), and therefore be inclined to choose subjects that return more social capital. In line with this interpretation, Ma (2009) shows that in a US sample, when accounting for prior attainment, lower SES students were more likely to study technical and business majors.

There are also large gender differences in uptake of STEM subjects throughout students' academic careers, and these disparities seem to grow larger over time, with only 19% of jobs in scientific sectors in the UK held by women (Kirkup et al., 2010). HESA statistics show that in 2013–2014, female students made up 48.3% of STEM undergraduates compared with 56.2% of students overall, and in engineering and technology subjects less than 10% of students were female (Equality Challenge Unit, 2014). For A levels, female students are less likely to study maths, physics and chemistry than male students, and more likely to study biology (Joint Council for Qualifications, 2014).

Unlike inequality in participation by students' family background, prior attainment cannot explain disparities by gender. There is a wealth of research considering difference in ability as the cause of gender disparities, however, this has been largely dismissed (Linn & Hyde, 1989) and it is widely accepted that in general, women and men are similar in abilities (Hyde, 2005). After conditioning on attainment, gender remains the largest predictor of uptake of maths at university (Noyes, 2009). In the UK, girls perform better in school than boys across most subjects, however, attainment is most similar for maths and science subjects. It could be that girls are less likely to choose STEM subjects because they achieve higher grades in other subjects, and therefore have more choice. Wang et al. (2013) show that students in a US college with high maths and verbal test scores were less likely to be working in STEM fields than those with high maths scores and average verbal scores. In consideration of these findings, the study presented considers the relationship between students' grades in maths, science and English individually, and whether English ability has a negative association with uptake.

The relationship between ethnicity and participation in particular subjects is complex, and strongly intertwined with family background, gender and prior attainment in the UK. In terms of academic capabilities, Strand (2007) studied Next Steps to understand the extent of differences in student attainment by ethnicity, showing that Pakistani, Bangladeshi, Black Caribbean and Black African students score lower in KS2 and KS3 examinations than their White British peers. When controlling for family background, most of these disparities were significantly reduced, however, Black Caribbean students continued to perform worse than expected. Differences in attainment generally even out by GCSE exams, with Black and Minority Ethnicity (BME) students having progressed at a faster rate than their White peers (Strand, 2014a).

Disparities in subject choice do not follow predicted patterns, given the relationship between attainment, family background and uptake of STEM subjects. Previous research looking across characteristics and using the Youth Cohort Study (YCS), the Labour Force Survey (LFS) and the Higher Education Statistics Agency (HESA) statistics showed that Chinese and Indian students were most likely to participate in Science, Engineering and Technology (SET) occupations, whilst African and Caribbean students, and Bangladeshi girls, were notably under-represented (Jones & Elias, 2005). The most recent data from HESA shows that overall, there is much higher ethnic diversity amongst STEM and other high-return university subjects (Equality Challenge Unit, 2014). For A-level choices, Black Caribbean students are least likely to study STEM subjects given their prior attainment, and White British students have particularly low uptake of maths (Boaler et al., 2011). It is likely that BME participation in STEM subjects will increase when taking into account students' prior attainment.

The reasons behind the increased uptake of STEM subjects amongst BME students are unclear. Research into biases in education point to numerous institutional disadvantages, particularly for black students. For example, there are particularly low representations of black individuals in science textbooks (Frost et al., 2005), and Black Caribbean students, given their attainment, are more likely to be put into lower ability groups (Strand, 2007). The latter is of particular concern in STEM subjects, where ability grouping is most often used (Kutnick et al., 2005). This may explain why black students appear least likely to study STEM subjects when compared with other minority ethnicity students, however, it does not explain why white students also appear to be under-represented. The relatively high ethnic diversity in STEM subjects is mirrored by a relative lack of diversity in arts and humanities subjects. It is possible that BME students are rejecting arts and humanities subjects, leading to higher proportions choosing STEM. Recent work has highlighted the issues of diversity in university curricula in the UK (Mirza & Joseph, 2013; Peters, 2015), especially considering the lack of representation of

BME individuals in philosophy, literature and history education.

Following a review of the literature in research detailing the relationship between ethnicity and attainment, Warikoo and Carter (2009) argue that the majority of studies rely on an additive model of student achievement, controlling for other student characteristics but not looking at differences in outcomes by combinations of characteristics. This chapter aims to address this by considering how student characteristics interact to influence their choices. For example, although gender and family background may both be negatively associated with choice, the magnitude and direction of the relationship between subject choice and student background may differ when we look within genders. There is a strong tradition in qualitative study of looking at the intersections between individuals' characteristics; at how individuals' experiences, given their characteristics, interact in more complex ways in producing disparities in outcomes (e.g. Crenshaw, 1989). Recent quantitative research looking into academic disparities has shown evidence for interactions (e.g. Dekkers et al., 2000; Kingdon & Cassen, 2010; Strand, 2014a).

5.1.2 Research questions

- What is the relationship between students' family background, gender and ethnicity with choice of STEM study at A level and university?
- Can disparities in uptake be explained by students' prior academic attainment?
- Do students' characteristics interact in determining choices?

This chapter proceeds as follows. The first section describes, under 'methodology', the data used for analysis, relevant variables and analytical strategy. The second section quantifies the proportions of students studying STEM at A level by students' gender, ethnicity and family background, and interactions between these characteristics. The third section considers HE subject choices. The fourth section concludes with a discussion of results and possible implications for policy and research.

5.2 Methodology

5.2.1 Data

I use Next Steps, previously the Longitudinal Study of Young People in England (LSYPE), a representative panel dataset including interviews, surveys and demographic information for young people and their parents or carers in England. The longitudinal nature of the data allowed me to compare student characteristics collected at age 14, with choices at age 18–19, eliminating the possibility that subject studied would influence the reporting of characteristics.

The study started in 2004, with the most recent wave of data collected in 2010. The sampling strategy for the study was twofold. Firstly, schools were sampled, with a focus on oversampling schools in deprived areas. Secondly, pupils within schools were sampled, with a focus on oversampling students from BME backgrounds. Owing to practical considerations, home-educated students, boarding students, students in schools with very small class sizes and students in the UK only for educational purposes were excluded from the study. Whilst the first four waves were collected via face-to-face interviews with young people and their parents or carers, the next three waves also employed telephone and Web-based survey methods. Full specifications of the sampling procedures employed in the study, and methods of data collection, can be found in the LSYPE user guide (Department for Education, 2011). The data has been linked with the National Pupil Database (NPD), giving detailed information on students' academic attainment across school years.

For key variables including the outcome (subject choice), ethnicity and gender, analysis is only carried out for individuals who gave valid responses. To retain adequate sample sizes, and avoid losing rich information on students who may have missing responses on a few variables, multiple imputation methods using chained equations were used for all other variables. It was not, however, considered meaningful to model students' ethnicity and gender based on other variables in the dataset. A total of 8494 students participated in Wave 1 and Wave 7 data collection (from which I draw my data), of which 4165 students had studied A levels and 4172 students were studying in HE, and gave valid responses for subject studied. Three students refused to report ethnicity, and a further 34 students from the A-level sample, and 37 from the degree sample, did not report sex. The final sample, therefore, was 4128 students studying A levels and 4132 students studying in HE. Table 5.1 further illustrates how the final samples were reached.

Table 5.1: Final sample size for this study compared with initial sample

Number of students	A level sample	Degree sample
Participated in Waves 1 & 7	8494	8494
Studied A levels, or at university, and reported subject choices	4165	4172
Reported subject choices, gender and ethnicity	4128	4132

In consideration of issues relating to attrition, weights provided and calculated by the UK data services (Department for Education, 2011) have been used for analysis. Weights for final analysis took into account the probability of students being in the initial sample (design weights) and the probability of response based on key variables (estimated through logistic regression methods). For Wave 7, variables associated with attrition included: gender, ethnic group, housing tenure, interview month, HE application status, and some behavioural traits. The purpose of using weights is to ensure that the sample remains representative of the population, and reduce the probability of bias due to differences in response rates. It is acknowledged that calculating weights is a complex process for longitudinal data, and that weights can only be applied based on students observed, and not unobserved characteristics. It is possible that there are unobserved characteristics, such as motivation, which may be associated with attrition, student characteristics and subject choice.

5.2.2 Key variables

Subject choice

Students' choice of 'at least one STEM A level', compared with studying no STEM subjects at A level, was modelled as a binary choice. STEM subjects at A level included maths, further maths, physics, chemistry and biology. Students in England typically study between three and four A levels, so their A-level choices may tell us less than HE choices about their future outcomes and careers. There remains a considerable financial return, however, to the study of STEM A levels, independent of HE subject choice [i.e. for maths A level, see Dolton and Vignoles (2002)]. Furthermore, a STEM university course will typically require at least one STEM subject studied at post-compulsory level (and usually two or more) for entry.

Students' subject choices at university were modelled as a categorical choice with three levels: STEM subjects; arts and humanities subjects; Social Sciences, Law and Business & administrative (SLB) subjects. STEM subjects in HE included: medicine and dentistry; subjects allied to medicine; biological sciences; veterinary sciences, agriculture and related; physical sciences; mathematical and computer sciences and engineering and technologies. 38.4% of students studied a STEM subject. All subjects considered under the broad umbrella of science were included in the STEM category during analysis, following research into STEM uptake also including biological and medical science (e.g. Botcherby & Buckner, 2012; Equality Challenge Unit, 2014). Whilst it is acknowledged that the largest gender disparities in uptake occur in physical sciences, and for biological and medical sciences this disparity isn't as large (see Boaler et al., 2011; Equality Challenge Unit, 2014), there remain large disparities in uptake of medicine and biological science by students' ethnicity and family background (van de Werfhorst et al., 2002; Equality Challenge Unit, 2014). Furthermore, it is clearly of policy interest to increase uptake of medical and biological sciences.

Walker and Zhu (2011) identified another group of subjects offering high returns to students following graduation: LEM (Law, Economics and Management). Because students' subject choices are grouped in Next Steps, students studying economics and management could not be identified individually. Instead, I included an indicator for students studying social studies (including economics), law and business & administrative studies, making up 29.9% of students. Remaining subject choices included: architecture, building and planning; linguistics, European languages; Eastern literature; history and philosophy; creative arts; education.

Table 5.2: Subjects included in groupings

Subject Choice	Subjects include
Arts and humanities	Architecture Building and planning Linguistics European language Eastern Literature History and philosophy Creative arts Education
STEM	Medicine and dentistry Subjects allied to medicine Biological sciences Veterinary sciences Agriculture and related Physical sciences Mathematical and computer sciences Engineering and technologies
SLB	Social sciences Law Business & administration studies

Family background

For initial analysis considering which family background indicators explain variation in subject choice, mothers' and fathers' highest academic qualification (degree and higher, A level and some HE, GCSEs and below), parents' NS-SEC occupational class (secretarial, intermediate, working class, long-term unemployed)[1] and students' gross family income[2] were included in all models.

Following prior research into family background differences in academic outcomes (e.g. Chowdry et al., 2011), an individual score was computed for each student to determine their socio-economic position (SEP) based on the following variables: how well the household is managing on finances; highest qualification of parents (whichever was highest); family's NS-SEC class and household tenure. I use polychoric principal components analysis (PCA) to identify a factor score and rank for each student. Although PCA is typically only appropriate for continuous variables, polychoric PCA has been shown to be an appropriate method for combining ordinal variables (see Kolenikov & Angeles, 2004). For the A-level and HE sample, the PCA factor explains 66% and 64%, respectively, of the variation in these indicators. In contrast to much prior research, 'eligibility' for FSM status was not used as a measure of economic status. Hobbs and Vignoles (2007) explain that generally, FSM eligibility is a poor proxy for student deprivation, and richer information is included on students' family income and other family background measures.

An indicator for whether students attended an independent school, or not, was included in the models. This follows research suggesting that independent-school students are more likely to study STEM and traditional subjects (e.g. CaSE, 2014). It is important to note that in Next Steps, independently educated students are underrepresented; 3.4% of students in the initial sample were independently educated compared with around 7% across England.

Attainment

Students' attainment was taken from NPD records; students' capped GCSE scores and individual scores in KS2 maths, science and English were included in the analysis. When splitting students into two attainment groups, above median attainment or below, large differences in participation by attainment are observed. Table 5.3 compares descriptive proportions of students in the high-attaining half of students by subject group. Students who study at least one STEM A level are more likely to be high achieving on a wide range of subjects. The largest difference is in GCSE scores, where 74% of students taking a STEM A level achieved above median scores. In line with A-level choices, students studying STEM subjects in HE are more likely to have higher scores across all indicators of attainment, except KS2 English, and those studying SLB have the lowest scores on average on all indicators except KS2 maths.

Table 5.3: Proportions of students scoring above average scores (compared to other cohort members) participating in each degree subject group, and for those taking at least 1 STEM subject at A-level

Subject	Take at least 1 STEM A-level	STEM Degree	SLB Degree	A&H Degree
High GCSE score	73.8%	60.2%	44.2%	48.2%
Above average KS2 Math score	69.8%	58.1%	45.3%	44.1%
Above average KS2 Science score	68.2%	61.7%	44.0%	52.4%
Above average KS2 English score	64.3%	57.2%	50.0%	58.1%

5.3 Analytical strategy

I first present raw descriptive statistics for students' choice of STEM A level, and of STEM and SLB subjects in HE, comparing proportions of students choosing each group of subjects by ethnicity and family background across genders. To understand which characteristics are most important in explaining students' subject choices, and how students' family background, gender and ethnicity interact in determining choice, I use logistic regression models. Regression methods identify the unique associations of each predictor variable with students' choices, thus allowing identification of which student characteristics explain the largest proportion of variance in choice, whilst other predictors are held constant.

Models are built up in three stages. Model 1 predicts students' subject choices based on their characteristics only. For A-level choices this is choice of at least one STEM A level compared with no STEM A levels. For degree subject choice, this is choice of STEM, SLB or arts and humanities subjects. Model 2 controls for prior attainment across subjects, and model 3 includes interaction terms[7]. For degree choices an additional fourth model is run, which also includes indicators for whether students studied STEM subjects at A level, to assess whether associations between student characteristics and degree choices are significant over and above their relationship with A-level choices.

The motivation for including all characteristics in the first model, rather than looking at raw proportions, is that student characteristics are strongly correlated. For example, students' SEB and ethnicity are strongly intertwined; the Labour Force Survey 2004 and the Pupil Leave School Census 2002 showed strikingly large differences in proportions of students claiming FSM (Bhattacharyya et al., 2003) or in relative income poverty (Kenway & Palmer, 2007). For this reason, it is likely that models not taking account of both student characteristics will under or over-estimate the diversity of uptake of STEM subjects. In the samples used for analysis, there are large differences in students' family background by their ethnicity. Table 5.4 outlines the proportions of students claiming FSM by ethnicity, which broadly reflect the proportions reported by Bhattacharyya et al. (2003). Students' attainment is also related to characteristics; students from lower SEBs especially are more likely to have lower levels of prior attainment, so it would be expected

[7] Versions of regressions were run using sub-group analysis rather than interaction terms, shown in Appendix B1.

that some of the differences in subject choice (especially choice of STEM subjects, which are considered 'harder' than other subjects) would reduce when accounting for attainment.

Table 5.4: Differences in proportions of students' claiming FSM by ethnicity[8]

Ethnicity	Unweighted count	Proportion claiming FSM
White British	2589	3.0%
Mixed	183	10.9%
Indian	478	6.6%
Pakistani	257	30.8%
Bangladeshi	232	56.4%
Black Caribbean	110	13.6%
Black African	154	28.3%
Other	132	20.7%

5.4 How do student characteristics interact in determining A-level subject choice?

Students typically study between three and four A levels, and given university entrance requirements it is unlikely that students who do not study at least one STEM A level will study a STEM subject at university. Proportions of female and male students from each ethnic group studying STEM A levels in the Next Steps sample are shown in Figure 5.1. As predicted, male students are more likely to study at least one STEM A level. Overall, Indian, Pakistani and 'other ethnicity' students are more likely to study STEM A levels than students from other ethnicities. White, Black African and Black Caribbean students have particularly low levels of relative uptake. There appear to be gender differences in uptake across the majority of ethnicities, with the exception of mixed ethnicity and Black Caribbean students, where there are no gender differences. Female students of mixed ethnicity and Black Caribbean ethnicity are more likely to study STEM A levels than white female students, whereas Black Caribbean male students are less likely to study STEM than white male students. For Bangladeshi students there is a particularly large gender disparity in proportions of students studying STEM, with just over 20% of young Bangladeshi women choosing STEM subjects at A level compared with over 50% of young Bangladeshi men.

[8] Taken from the sample attending university, results were similar for the A-level sample.

Figure 5.1: Raw proportions of students who completed at least one STEM A-level by ethnicity

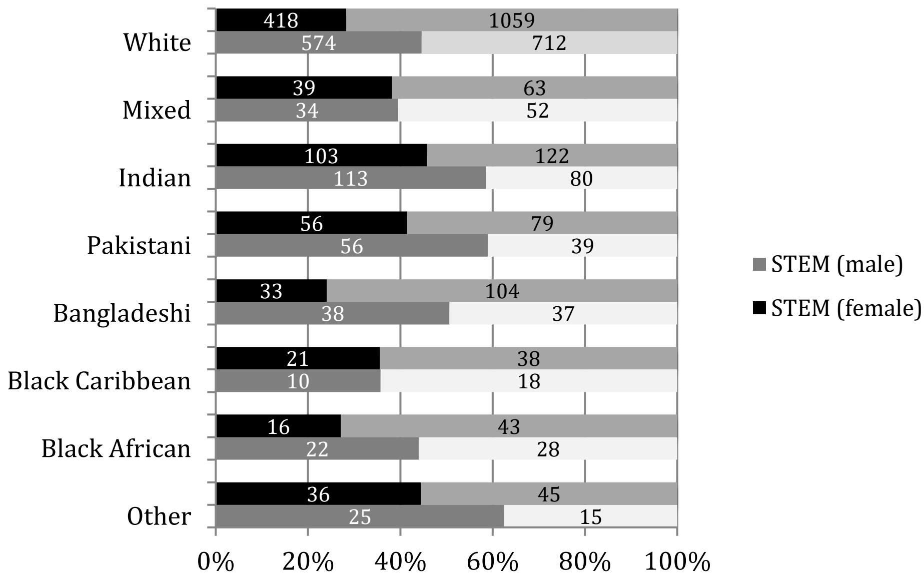


Table 5.5 illustrates the relationship between students' family background, gender and subject choice. Male students taking at least one STEM A level are more likely to be in higher income bands, and all students choosing STEM A levels are more likely to have parents with higher educational achievements and in higher occupational classes than students who were not studying any STEM subjects. They are also more likely to be attending independent schools, and to be in the highest SEP group.

Table 5.5: Family background characteristics of female (male) students completing at least 1 STEM A-level[9]

Subject	Take at least 1 STEM subject	No STEM subject
Median Income band	£28600 - £31200 (£31200 -£33800)	£28600 - £31200 (£28600 -£31200)
Mother has Degree or higher	20.2% (23.4%)	14.0% (16.3%)
Father has Degree or higher	25.4% (26.3%)	13.4% (17.1%)
Household has service class occupation	58.1% (57.7%)	50.0% (52.1%)
Independently educated	19% (15.6%)	11% (13.9%)
Highest SEP	36.7% (40.7%)	27.7% (29.3%)

5.4.1 Regression models of A-level subject choices

Logistic regression results of the relationship between students' characteristics and subject choices are shown in Table 5.6. The first model includes students' ethnicity, family background indicators and school type. The second model additionally includes students' prior academic attainment and the third model includes interaction terms. Figure 5.2 illustrates differences in students' odds of choosing at least one STEM A level by ethnicity, with the blue dots illustrating odds before conditioning on attainment (taken from model 1) and the purple dots illustrating odds after conditioning on attainment (taken from model 2). Differences in choices by ethnicity broadly reflect raw associations, however, Figure 5.2 shows that with the addition of prior attainment to the regressions, differences in uptake increase substantially. This suggests that the full extent of disparities in choice by ethnicity is suppressed by attainment differences, which influence choices in the opposite direction.

[9] Incomes are at 2003 prices, measured in wave 1 data collection. 'Service class' occupations includes parents in higher and lower managerial and professional occupations, and parent with 'at least some HE' includes parents with some HE, and those a Degree qualification or higher.

Table 5.6: Results of logistic regression of choice of STEM A-level, odds Ratios are shown with standard errors in parenthesis.

	Model 1		Model 2		Model 3	
Variables	OR	SE	OR	SE	OR	SE
Female	0.533***	(0.041)	0.499***	(0.050)	0.494***	(0.055)
Ethnicity (Ref: White)						
Mixed	1.389	(0.307)	1.468	(0.438)	1.056	(0.535)
Indian	2.570***	(0.344)	3.793***	(0.665)	3.747***	(1.011)
Pakistani	2.749***	(0.497)	5.260***	(1.155)	5.281***	(1.639)
Bangladeshi	1.424*	(0.274)	1.792**	(0.434)	3.283***	(1.215)
Black Caribbean	1.717*	(0.536)	3.877***	(1.243)	2.116	(0.968)
Black African	1.079	(0.255)	2.278***	(0.716)	1.867	(0.924)
Other	3.425***	(0.868)	3.914***	(1.161)	4.543***	(2.559)
Independent school	1.192	(0.165)	0.836	(0.179)	0.825	(0.176)
Mother Highest Qual (Ref: GCSE's or lower)						
Degree or Higher	1.172	(0.136)	0.676**	(0.103)	0.664***	(0.102)
A-levels or some HE	0.957	(0.092)	0.752**	(0.087)	0.751**	(0.087)
Mum not present	0.393**	(0.150)	0.441**	(0.166)	0.442**	(0.166)
Fathers Highest Qual (Ref: GCSE's or lower)						
Degree or Higher	2.016***	(0.252)	1.541***	(0.233)	1.715***	(0.289)
A-levels or some HE	1.216*	(0.129)	1.105	(0.144)	1.168	(0.157)
Dad not present	0.957	(0.131)	0.991	(0.156)	0.916	(0.159)
Social class (Ref: Working class)						
Managerial	1.244*	(0.146)	1.007	(0.142)	1.051	(0.156)
Intermediate	1.119	(0.145)	1.173	(0.175)	1.195	(0.181)
Unemployed	0.817	(0.209)	0.901	(0.287)	0.858	(0.296)
Income	1.002	(0.003)	0.999	(0.004)	1.000	(0.004)
Attainment						
GCSE			3.354***	(0.272)	3.371***	(0.274)
Ks2 Math			2.278***	(0.187)	2.283***	(0.189)
Ks2 Science			1.248***	(0.103)	1.253***	(0.103)
Ks2 English			0.537***	(0.039)	0.538***	(0.039)
*Female*SEP*					0.958	(0.098)
Ethnicity x SEP						
Mixed*SEP					0.715	(0.215)
Indian *SEP					0.813	(0.155)
Pakistani*SEP					1.002	(0.213)
Bangladeshi*SEP					0.691*	(0.139)
Black Caribbean*SEP					1.070	(0.248)
Black African*SEP					1.104	(0.248)
Other*SEP					1.008	(0.197)
Ethnicity x Sex						
Mixed*Female					1.818	(1.121)
Indian*Female					1.028	(0.346)
Pakistani*Female					0.946	(0.382)
Bangladeshi*Female					0.591	(0.264)
Black Caribbean*Female					2.037	(1.147)
Black African*Female					0.984	(0.609)
Other*Female					0.689	(0.418)
Constant	0.447***	(0.063)	0.506***	(0.082)	0.479***	(0.082)
Observations			4,128			

*** p<0.01, ** p<0.05, * p<0.1

Figure 5.2: Students odds of studying at least one STEM A level by their ethnicity[10]

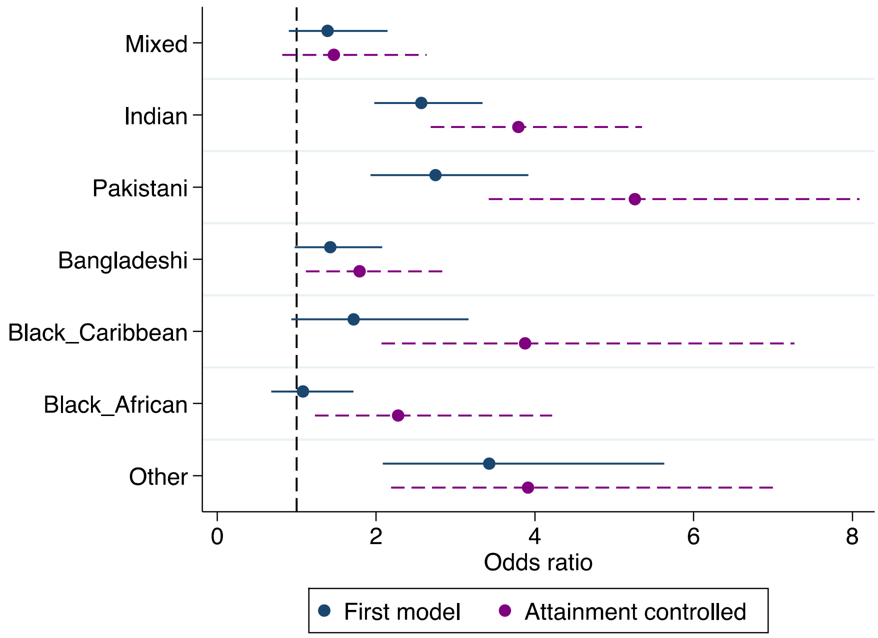
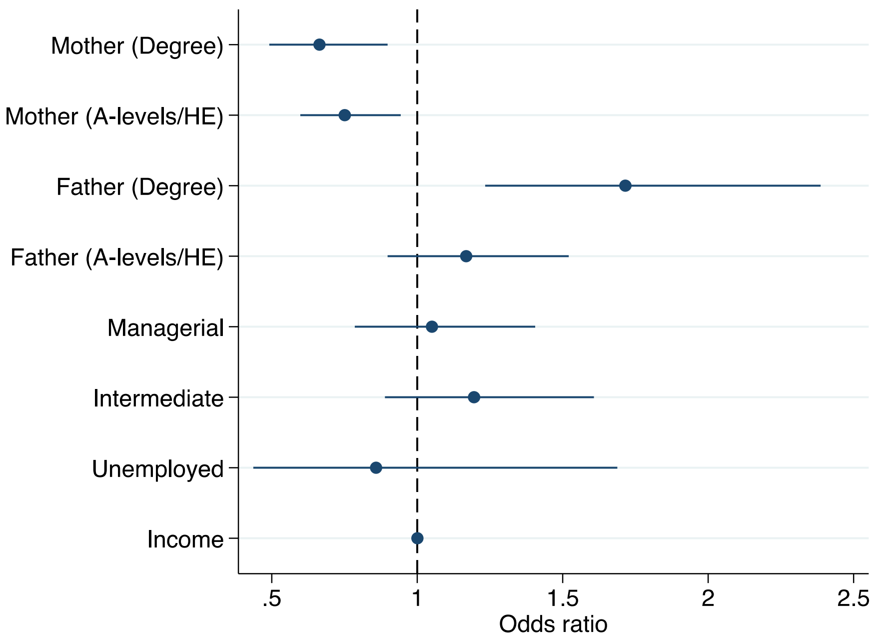


Figure 5.3: Students odds of studying at least one STEM A-level by family background characteristics[11]



[10] The reference category is White students
[11] The reference category is students from working class backgrounds whose parents do not have qualifications higher than GCSE level

One possible reason why BME students may be more likely to choose higher-return STEM subjects could be related to differences in parental and student attitudes and behaviours; BME groups generally have more favourable scores on these characteristics when considering outcomes (Strand, 2011). Whilst Strand found that an increase in these attitudes and behaviours does not lead to proportionately higher academic attainment, they could influence student choices.

Students' social class and parents' education are both uniquely related to choices. Students whose parents work in managerial occupations are more likely to study STEM than students with parents in working-class occupations. The relationship is, however, fully explained by prior attainment. Students from higher social classes are more likely to achieve higher grades, which in turn predicts participation in STEM A levels. Parents' education levels have differing associations with STEM study, which persist when conditioning on attainment. Students whose mothers have a degree are less likely to study STEM A levels, whilst students whose fathers have a degree are more likely to study STEM A levels. Figure 5.3 illustrates this relationship between students' family background and choice of STEM A levels with all student characteristics, attainment measures and interaction terms controlled, showing how the association between both parents' education and choices persists, whilst other background characteristics are no longer significantly associated with choices.

Compared with other family background characteristics, parental income and whether students attended independent school are not associated with participation in STEM. This suggests that students 'parents' education and social class, rather than differences in schooling drives relationships between type of school and participation.

Overall, students' prior attainment is positively associated with choice, with the exception of KS2 English attainment. This is in line with research by Wang et al. (2013), and suggests that students who do well in English are choosing to pursue other subjects. It is noted that due to the issue of multicollinearity, care should be taken when interpreting the odds ratios on attainment scores; scores are likely to be highly correlated and therefore exact values would change considerably with the addition or subtraction of indicators in the model. As it stands, we can only confidently ascertain direction of association and the cumulative effect of attainment indicators on other associations.

Overall, there are few interactions between student characteristics and A-level choices, the only exception being that more advantaged Bangladeshi students are less likely to pursue STEM subjects at A level.

5.5 How do student characteristics interact in determining HE subject choice?

There are well-established differences in choice by students' gender; male students are more likely to study STEM subjects at university, whilst female students are more likely to study arts and humanities. In terms of ethnicity, HESA data covering students across the UK also reveals that overall, students from BME backgrounds are more likely to study STEM and SLB subjects and less likely to study arts and humanities subjects, although there is large heterogeneity between ethnic groups and subjects (Equality Challenge Unit, 2014). The Next Steps data also indicates that there are large differences in participation by students' gender and ethnicity, as shown in Figure 5.4. White students are least likely to study high-return SLB subjects, whilst Asian students are most highly represented, and this increase in uptake is mirrored by very low uptake of arts and humanities subjects. Black Caribbean and Black African students stand out as being particularly under-represented in STEM.

Table 5.7 illustrates the raw relationships between family background, gender and subject choice. There are small differences in average income of students in each subject group. Female students studying SLB subjects have the lowest median family incomes, whereas young men studying either SLB or arts and humanities subjects have the lowest family incomes. Students studying STEM and arts and humanities subjects are most likely to have parents with a degree or higher, and in service-class occupations, compared with students studying SLB subjects. In contrast, SLB subjects appear to attract the highest proportions of independently educated students. In considering students' SEP, SLB subjects stand out as having particularly low uptake amongst the most advantaged female students, whilst for male students, differences between groups are small.

Figure 5.4: Raw proportions of students studying STEM, SLB or other subjects at university by ethnicity and gender

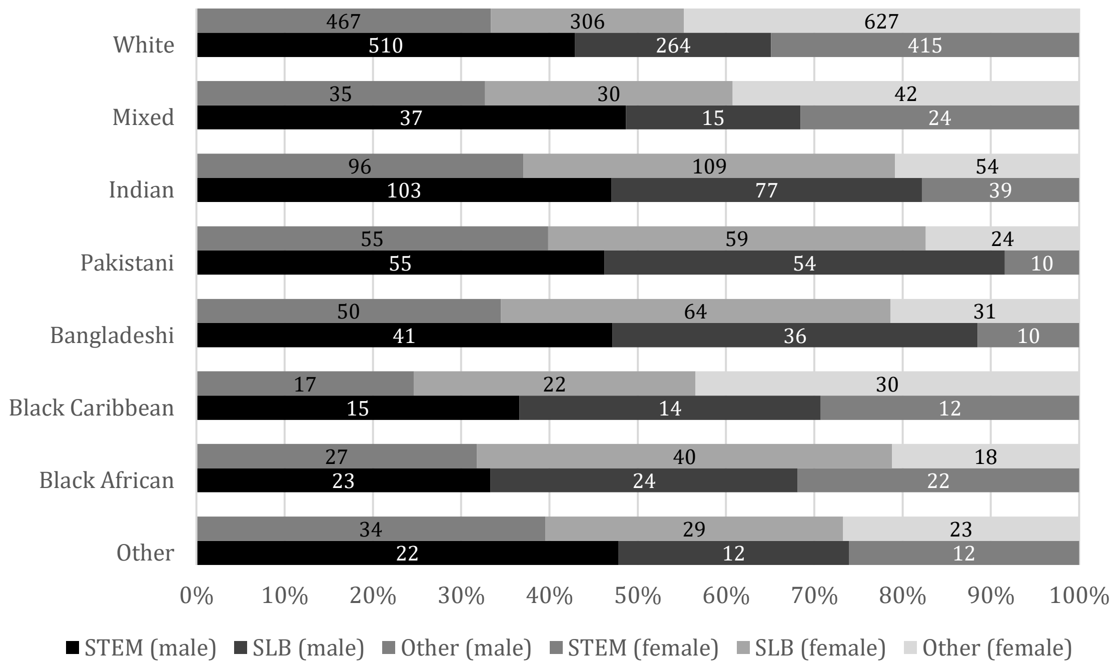


Table 5.7: Family background characteristics of female (male) students choosing STEM, SLB or other degree subjects[12]

Subject	STEM	SLB	A&H
Median Income band	£28600 - £31200 (£31200 - £33800)	£26000 - £28600 (£26000 - £28600)	£28600 - £31200 (£26000 - £28600)
Mother has Degree or higher	16.5% (19.2%)	9.6% (16.3%)	17.4% (22.8%)
Father has Degree or higher	20.5% (21.7%)	12.1% (18.8%)	17.2% (23.9%)
Managerial class	53.7% (55.4%)	46.4% (51%)	54.4% (54.5%)
Independently educated	1.6% (3.4%)	3.5% (4.9%)	3.1% (3.6%)
Highest SEP	32.5% (35.5%)	21.7% (30.4%)	32.1% (36.8%)

[12] Incomes are at 2003 prices, measured in wave 1 data collection. 'Service class' occupations includes parents in higher and lower managerial and professional occupations, and parent with 'at least some HE' includes parents with some HE, and those a Degree qualification or higher.

Table 5.8: Results of multinomial logistic regression of degree choice

	Arts and humanities								Social sciences, Business and Law							
	Model 1		Model 2		Model 3		Model 4		Model 1		Model 2		Model 3		Model 4	
VARIABLES	RRR	SE	RRR	SE	RRR	SE	RRR	SE	RRR	SE	RRR	SE	RRR	SE	RRR	SE
Female	1.538***	(0.131)	1.330***	(0.122)	1.393***	(0.146)	1.065	(0.124)	1.211**	(0.112)	1.194*	(0.120)	1.191	(0.146)	0.917	(0.121)
Ethnicity (Ref: White)																
Mixed	0.978	(0.229)	0.943	(0.226)	1.125	(0.393)	1.149	(0.439)	1.276	(0.336)	1.261	(0.342)	0.836	(0.356)	0.860	(0.385)
Indian	0.490***	(0.089)	0.451***	(0.084)	0.509**	(0.152)	0.888	(0.339)	1.497***	(0.214)	1.417**	(0.209)	1.345	(0.289)	2.272***	(0.653)
Pakistani	0.235***	(0.058)	0.202***	(0.049)	0.186***	(0.075)	0.326**	(0.142)	1.426*	(0.275)	1.304	(0.258)	1.556*	(0.387)	2.713***	(0.695)
Bangladeshi	0.406***	(0.110)	0.369***	(0.100)	0.234***	(0.107)	0.345**	(0.174)	1.676***	(0.332)	1.648**	(0.332)	1.377	(0.473)	2.000*	(0.733)
Black Caribbean	0.928	(0.314)	0.667	(0.263)	0.675	(0.310)	0.662	(0.310)	1.304	(0.443)	1.124	(0.387)	1.311	(0.584)	1.250	(0.576)
Black African	0.671	(0.170)	0.499**	(0.135)	0.925	(0.347)	0.878	(0.366)	1.952***	(0.464)	1.683**	(0.423)	1.899*	(0.729)	1.842	(0.721)
Other	0.450***	(0.130)	0.450***	(0.131)	0.624	(0.267)	1.407	(0.554)	0.794	(0.210)	0.839	(0.221)	0.758	(0.318)	1.594	(0.714)
Independent school	1.472	(0.376)	1.458	(0.382)	1.438	(0.369)	1.288	(0.337)	1.781**	(0.455)	1.745**	(0.449)	1.754**	(0.454)	1.586*	(0.412)
Mother Highest Qualification (Ref: GCSE's or lower)																
Degree or Higher	1.103	(0.142)	1.286*	(0.176)	1.294*	(0.178)	1.407**	(0.217)	0.785	(0.120)	0.882	(0.138)	0.881	(0.139)	0.955	(0.167)
A-levels or some HE	0.947	(0.101)	1.010	(0.110)	1.012	(0.111)	0.929	(0.111)	0.995	(0.117)	1.035	(0.123)	1.038	(0.124)	0.960	(0.125)
Mum not present	0.850	(0.340)	0.806	(0.316)	0.782	(0.307)	0.672	(0.272)	1.041	(0.401)	1.008	(0.393)	1.039	(0.409)	0.932	(0.369)
Fathers Highest Qualification (Ref: GCSE's or lower)																
Degree or Higher	0.875	(0.123)	1.011	(0.148)	0.981	(0.186)	1.320	(0.269)	0.702**	(0.115)	0.773	(0.128)	0.779	(0.159)	1.024	(0.222)
A-levels/ some HE	1.040	(0.129)	1.115	(0.143)	1.104	(0.165)	1.247	(0.203)	0.992	(0.136)	1.035	(0.144)	1.048	(0.167)	1.189	(0.202)
Dad not present	1.090	(0.155)	1.088	(0.162)	1.140	(0.209)	1.118	(0.220)	1.062	(0.164)	1.053	(0.164)	1.020	(0.195)	0.983	(0.204)
Social class (Ref: Working class)																
Managerial	0.968	(0.128)	1.010	(0.138)	1.008	(0.140)	1.019	(0.156)	0.921	(0.131)	0.962	(0.138)	0.975	(0.142)	0.991	(0.158)
Intermediate	1.226	(0.177)	1.244	(0.184)	1.240	(0.185)	1.388**	(0.230)	1.124	(0.170)	1.138	(0.173)	1.145	(0.176)	1.279	(0.212)
Unemployed	1.274	(0.377)	1.267	(0.383)	1.315	(0.410)	1.293	(0.417)	0.855	(0.237)	0.850	(0.234)	0.845	(0.246)	0.833	(0.253)
Income	1.000	(0.004)	1.002	(0.004)	1.001	(0.004)	1.006	(0.004)	1.001	(0.004)	1.002	(0.005)	1.002	(0.005)	1.006	(0.005)
Attainment																
GCSE			0.779***	(0.061)	0.775***	(0.061)	1.174**	(0.093)			0.732***	(0.058)	0.728***	(0.058)	1.089	(0.094)
Ks2 Math			0.677***	(0.053)	0.678***	(0.054)	0.877	(0.076)			1.039	(0.092)	1.041	(0.093)	1.314***	(0.125)
Ks2 Science			0.886	(0.070)	0.888	(0.070)	0.995	(0.088)			0.787***	(0.067)	0.786***	(0.067)	0.874	(0.081)
Ks2 English			1.404***	(0.119)	1.405***	(0.120)	1.153*	(0.097)			1.296***	(0.111)	1.304***	(0.111)	1.094	(0.097)
Female*SEP					1.004	(0.104)	1.021	(0.112)					0.790**	(0.087)	0.798*	(0.093)
Ethnicity x SEP																
Mixed*SEP					1.276	(0.270)	1.517*	(0.326)					1.185	(0.252)	1.387	(0.289)
Indian *SEP					0.638**	(0.134)	0.666	(0.167)					0.880	(0.144)	0.917	(0.165)
Pakistani*SEP					1.072	(0.233)	1.072	(0.239)					1.203	(0.238)	1.225	(0.236)
Bangladeshi*SEP					0.985	(0.214)	1.079	(0.259)					0.814	(0.177)	0.867	(0.196)
Black Caribbean*SEP					0.998	(0.295)	0.980	(0.295)					0.906	(0.238)	0.878	(0.251)

Black African*SEP					0.976	(0.209)	0.916	(0.218)					1.184	(0.246)	1.125	(0.255)
Other*SEP					1.150	(0.273)	1.386	(0.349)					0.917	(0.228)	1.084	(0.309)
Ethnicity x Female																
Mixed*Female					0.828	(0.391)	0.965	(0.478)					2.047	(1.133)	2.369	(1.331)
Indian*Female					0.760	(0.271)	0.862	(0.363)					1.111	(0.311)	1.260	(0.418)
Pakistani*Female					1.168	(0.577)	1.295	(0.666)					0.813	(0.293)	0.883	(0.308)
Bangladeshi*Female					1.949	(0.976)	2.105	(1.121)					0.899	(0.350)	0.941	(0.365)
Black Caribbean*Female					0.976	(0.694)	1.382	(0.979)					0.637	(0.409)	0.859	(0.566)
Black African*Female					0.290**	(0.152)	0.360*	(0.202)					0.985	(0.475)	1.203	(0.627)
Other*Female					0.607	(0.338)	0.521	(0.289)					1.041	(0.575)	0.914	(0.557)
Studied 1 STEM A level							0.268***	(0.037)							0.343***	(0.049)
Studied 2 or more STEM A levels							0.034***	(0.006)							0.052***	(0.009)
Constant	0.845	(0.132)	0.800	(0.126)	0.787	(0.129)	1.462**	(0.271)	0.618***	(0.106)	0.579***	(0.102)	0.568***	(0.107)	1.004	**(0.206)**
Observations	4,135		4,135		4,135		4,135		4,135		4,135		4,135			4,135

*** p<0.01, ** p<0.05, * p<0.1

5.5.1 Regression models of HE subject choices

Table 5.8 presents results from multinomial logistic regressions of the relationship between subject studied and students' characteristics. Like A-level choices, regression models were built up in stages, with the first model including only student characteristics and school type, the second model conditioning on attainment and the third model including interaction terms. A fourth model is run, including indicators for whether students studied STEM at A level.

Differences in choice by ethnicity are strikingly large. The first model shows that, even after accounting for family background, students from BME backgrounds are less likely to study arts and humanities subjects, and more likely to study SLB subjects, than STEM subjects. Black Caribbean students and students of mixed ethnicity, however, are most similar to white students in their choices, and are no more likely to study STEM (see Figure 5.5).

In line with raw associations and prior research, differences in uptake of STEM and other subjects are observed by students' family background (Gorard & See, 2009; The Royal Society, 2008). Whilst social class and family income are not significantly associated with choices, parental education (particularly mothers' highest qualification) is. Students whose mothers have a degree are more likely to study arts and humanities than STEM subjects, even when prior attainment differences are taken into account. Students whose fathers have a degree are more likely to study STEM than SLB subjects, however, this relationship is fully explained by attainment differences (see Figure 5.6).

It might be expected, given that STEM and SLB subjects offer higher financial returns, that family income would be associated with choices, for example students from higher-income families may be more concerned with financial returns after study (e.g. Davies et al., 2013). Alternatively, students from lower-income families may be more inclined to avoid more risky subjects when considering outcomes (e.g. Breen & Goldthorpe, 1997) and choose 'easier' subjects. Despite this, and raw statistics indicate otherwise, when taking account of other student characteristics, family income is not related to subject studied. In terms of schooling, there is an indication that independently educated students, all else held equal, are more likely to study high-return SLB subjects over STEM subjects.

Interactions are observed between students' social background and gender, and between ethnicity and gender. As students' SEP increases, young women are less likely to choose SLB subjects and more likely to choose to study STEM subjects. This suggests that young women from more deprived backgrounds may be particularly vulnerable to factors driving students away from STEM. Black African female students, however, are much more likely to choose STEM over arts and humanities.

Model 4 shows that when including indicators for whether students studied one STEM A level or two or more STEM A levels, results are largely similar. Students studying STEM at A level were considerably more likely to study STEM subjects at degree over both arts and humanities, and SLB subjects. Taking account of A-level choices did affect some ethnic differences in participation, for example Indian students who studied STEM at A level were not significantly more likely to choose STEM over arts and humanities than white students who also studied STEM A levels. In contrast, when accounting for A-level choices, Pakistani and (to a lesser extent) Bangladeshi students were more likely to choose SLB over STEM compared with white students. The social background disparities persisted and increased somewhat, with students whose mothers had a degree remaining more likely to study arts and humanities than students with lower levels of education. The interactions between gender and SEP, and between Black African ethnicity and gender in determining uptake, also persisted.

Figure 5.5: Students odds of studying Arts and Humanities, or SLB subjects over STEM subjects at university by ethnicity[13]

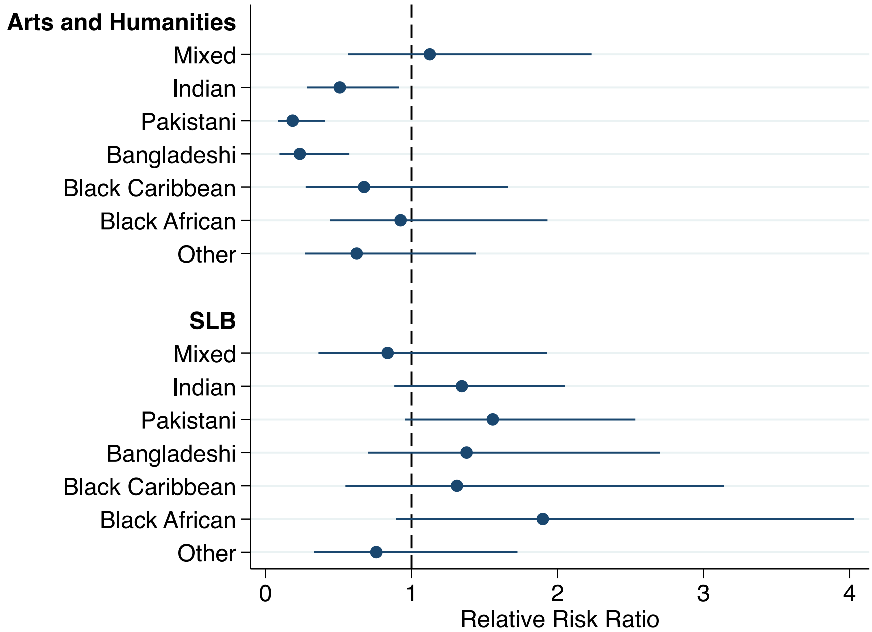
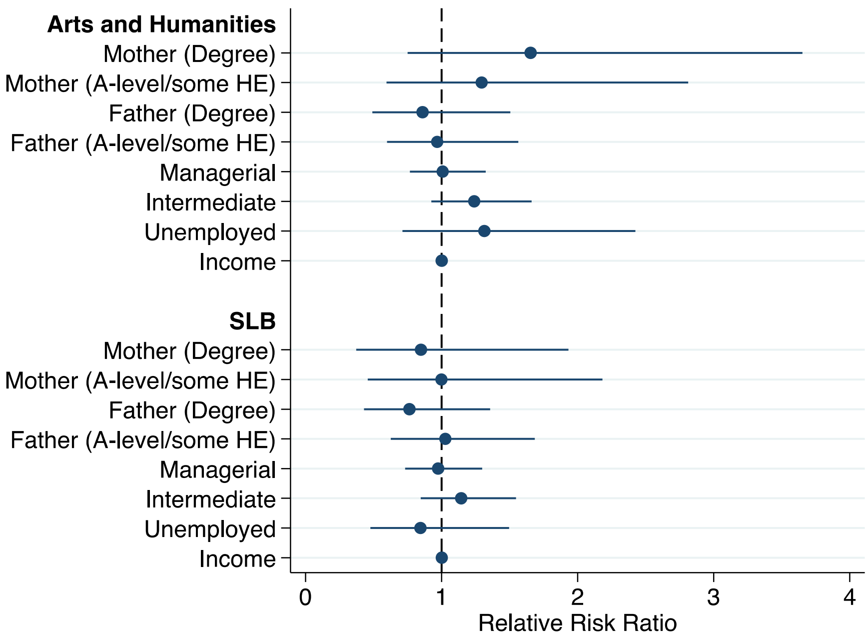


Figure 5.6: Students odds of studying Arts and Humanities, or SLB subjects over STEM subjects at university by family background

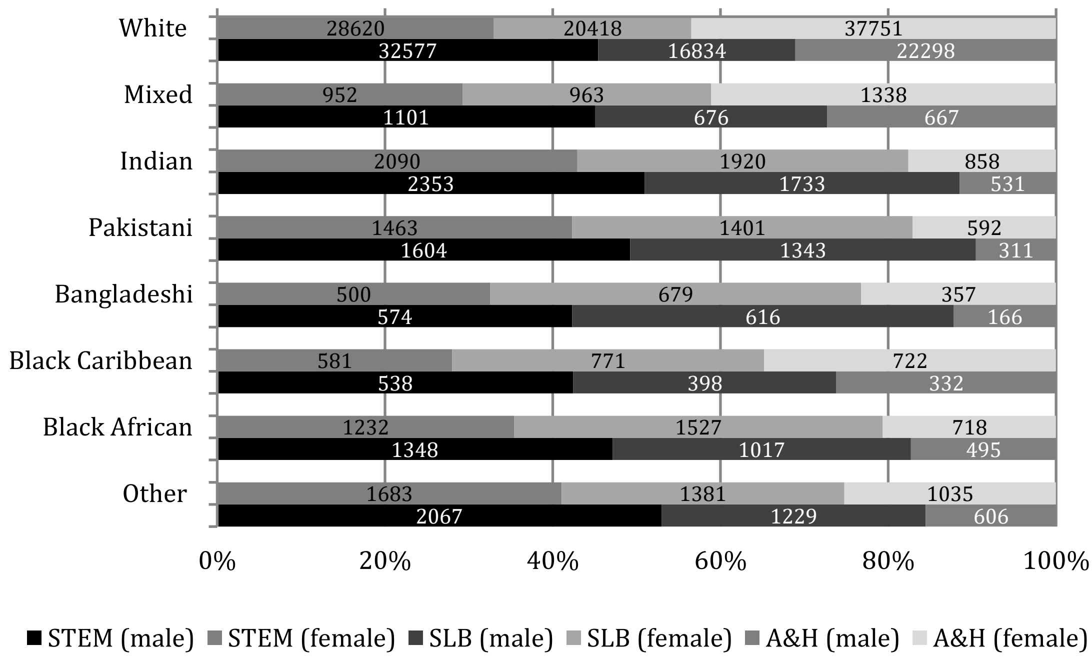


[13] The reference category is White students, and results from the final model are shown (including attainment and interaction effects).

5.5.1 Validating results with HESA-NPD linked data

A number of relationships were identified in this chapter between students' characteristics and their subject choices at university. Access to linked data from HESA and the NPD was granted at the later stages of writing this thesis, and included information on all students who attended higher education in the UK in the year 2009/2010, with information on their course of choice, ethnicity, gender and family background. The measures of family background available are less detailed than those available in Next Steps, however the HESA dataset includes information on whether students had a parent who attended university or higher education. It should be noted that ethnicity information was missing, refused or unknown for 34% of students, and parent's education was missing for 25% of students.

Figure 5.7: Student ethnicity, gender and subject choices in higher education from the HESA-NPD dataset (Total N – 200,966)



Results using the HESA-NPD data shown in figure 5.7 largely reflect those found in the Next Steps sample (shown in figure 5.4). For students from all ethnic groups, young men are more likely to be studying STEM then young women. Indian and Pakistani students are most highly represented in STEM subjects, and have lower uptake of art and humanities subjects.

Table 5.9: Proportions of students whose parents have a degree or HE qualification by gender and subject studied (Total N – 228,097)

Subject - % (N)	STEM	SLB	A&H
Female	53.7% (23,668)	49.1% (16,363)	54.7% (26,324)
Male	56.4% (26,615)	53.7% (14,567)	57.2% (16,061)

Table 5.9 shows the proportion of students studying each group of subjects who have parents with a degree or some higher education, for male and female students separately. Over half (54.2%) of students attending university had highly educated parents. Young women attending university had less educated parents, perhaps related to the fact that more young women attend university overall. Students studying SLB subjects were least likely to have highly educated parents.

Finally, a regression was run similar to that shown in model 3 in table 5.8. The regression model included students' gender, ethnicity, and family background, however did not include students prior attainment in school. Students with missing data on any of these characteristics were not included in the model.

Table 5.10: Multinomial logistic regression results predicting subject choice based on student characteristics

	Arts and Humanities		SLB	
Variables	RRR	SE	RRR	SE
Female	1.925***	(0.034)	1.407***	(0.025)
Ethnicity (Ref: White)				
Mixed	0.983	(0.036)	1.342***	(0.051)
Indian	0.297***	(0.011)	1.290***	(0.035)
Pakistani	0.305***	(0.014)	1.373***	(0.044)
Bangladeshi	0.486***	(0.032)	1.799***	(0.090)
Black Caribbean	0.921	(0.048)	1.706***	(0.086)
Black African	0.480***	(0.020)	1.690***	(0.057)
Other	0.443***	(0.016)	1.118***	(0.035)
Parent has a degree/ HE	0.988	(0.018)	0.875***	(0.016)
Female*Parent has a degree/ HE	1.025	(0.025)	0.939**	(0.024)
Constant	0.658***	(0.009)	0.553***	(0.008)
Observations	157,173			

*** p<0.01, ** p<0.05, * p<0.1

In line with results from table 5.8, women were more likely to study both arts and humanities subjects, and SLB subjects, over STEM subjects compared to men. Indian, Pakistani, Bangladeshi and black African students were more likely to choose STEM over arts and humanities subjects than white students, and students from all ethnic groups were more likely to choose SLB subjects over STEM than white students. Students with highly educated parents were more likely choose STEM over SLB compared to students whose parents had lower levels of education, and the interaction term suggests this is particularly the case for young women. The direction of this result was confirmed with subgroup analysis; with regressions run for students whose parents had low and high education separately, shown in B2.

5.6 Discussion

This chapter aimed to describe disparities in students' subject choices by their family background, ethnicity and gender, and to unpick the more complex relationships between these characteristics. I focused specifically on uptake of STEM subjects at A level and HE because these subjects have high levels of disparity in uptake across student characteristics, as well as numerous benefits of study to both individuals and society. For HE choices this was compared with uptake of two other groups of subjects: SLB subjects, which offer higher returns on graduation to individuals, and arts and humanities subjects. Although research into educational achievement disparities has started to look at how student characteristics interact to produce outcomes, rather than simply how they additively lead to deficit in attainment, studies of students' subject choices have not yet considered more complex models. The study addressed this by looking at whether family background could explain disparities in uptake by students' ethnicity, and whether patterns of choice differed for male and female students, or across socio-economic groups.

The findings complement a growing literature profiling disparities in uptake of STEM subjects (e.g. Gorard & See, 2009; Boaler et al., 2011; Botcherby & Buckner, 2012). In the Next Steps sample, students of almost all minority ethnic groups were more likely to study STEM and SLB subjects given family background, and this association increased when taking account of their prior attainment. Although generally there were similar patterns of uptake by students' ethnicity across genders, the interaction between Black African ethnicity and gender suggests that Black African women are more likely to study STEM than arts and humanities. This is in contrast to raw data suggesting that Black African and Caribbean students are less likely to study STEM subjects when family background is not accounted for (Boaler et al., 2011). It is possible that the underlying reasons for these differences, whether driven by cultural differences or biases (institutional or individual), are affecting young women and men differently. The findings offer additional evidence of the relative lack of ethnic diversity in arts and humanities subjects, where white students are disproportionately represented compared with all other ethnic groups. In terms of theories of relative risk aversion, given that there appear to be some additional barriers within HE and upon graduation for BME students, they may be making very rational choices to study subjects which have more secure prospects and higher financial returns. For example, research figures show that in the UK, minority ethnicity students are less likely to receive high degree classifications and are more likely to be unemployed after graduation (Runnymede Trust, 2014).

This chapter adds to the literature by considering a more comprehensive range of indicators for students' family background, including income, parents' education, occupational status and type of school attended. It appears that parental education, but not social class or financial resources, influence students' choices. Students studying STEM A levels are more likely to have fathers with a degree, and less likely to have mothers with a degree. At degree level, students whose mothers have a degree are most likely to study arts and humanities. It is possible that this relationship is related to the subject parents are educated in, and relative 'science capital' in the family (Archer et al., 2012). As mothers are more likely to have non-science degrees than fathers, they may influence their children to study other subjects. Because the Next Steps data does not include subjects studied by parents, this isn't something that could be explored further in the current study.

The interaction between students' family background and gender suggests that young women from more advantaged backgrounds are more likely to choose STEM subjects, whilst those from relatively deprived backgrounds are more likely to study SLB subjects, which—although they offer high individual returns—are not considered 'difficult' compared with STEM subjects. In accordance with the theory of relative risk aversion, more advantaged female students may be choosing more 'risky' high-return subjects compared with their less advantaged peers.

As with ethnicity, there isn't sufficient evidence that young women have an innate difference in ability to young men, and much research has profiled the many institutional biases that may push young women away from STEM subjects. STEM subjects are stereotypically seen as more ‘masculine’ domains, and in school, girls with the same academic attainment as boys are less likely to be rated as high achieving in maths by teachers (Campbell, 2015) and less likely to receive positive reinforcement from teachers (Mujtaba & Reiss, 2012), which may affect self-efficacy beliefs. Although the reasons are unclear, girls are less likely to be interested in science, and more likely to be interested in people, than boys (Collings & Smithers, 1984). What sets this work apart is the finding that disparities are not constant, but differ by students’ family circumstances. Given the institutional factors at play throughout students’ lives, it may be that the processes involved in overcoming stereotypes are also associated with students’ background. Students from lower SEPs may be more likely to feel constrained by their gender and to feel that they have less control over their future, which may in turn be related to uptake (e.g. Mau et al., 1995).

It could also be that students’ family background is related to parents’ attitudes and behaviours, which mediate the relationships observed. If mothers with higher education levels have more egalitarian views of gender roles (Crompton & Lyonette, 2005), these views may be transmitted to their children (Kulik, 2002; Antill et al., 2003) and thus directly or indirectly influence young women’s interests and values when choosing courses. Future research could focus specifically on whether student and parental attitudes and behaviours mediate the relationship between students’ characteristics, SEP and subject choices.

There are various strengths to the analysis presented. Based on observable characteristics, Next Steps data is generally representative of the population, and weights are applied where this is not the case. This is a recent sample, and students’ subject choices in 2008–2010 are analysed. Furthermore, I have included a rich set of student family background characteristics to draw evidence from, and the longitudinal nature of the dataset allows me to assess whether student circumstances at age 13–14 can predict later subject choices. Despite these strengths, there remain some limitations to the study. Although weights have been applied to ensure the data are representative, these could only be modelled on observed characteristics, and it is possible that there are some unobserved characteristics related to both non-participation and subject choice. In addition, the majority of indicators (with the exception of student attainment) are based on self-report from students and parents, which may lead to some measurement error. Recent policy changes, such as the increase in the student fees cap from 2012, may have an effect on students' subject choices; something that cannot be assessed in the current Next Steps cohort.

Chapter 6: What role do students' enjoyment and perception of ability play in social disparities in subject choices at university?

6.1 Introduction

Considerable research has outlined educational inequalities in the UK, and the mechanisms through which more advantaged families help their children to achieve higher levels of education (e.g. Blanden & Gregg, 2004; Blanden, Gregg, & Machin, 2005; Bukodi, Goldthorpe, Waller, & Kuha, 2015; Goldthorpe & Mills, 2008). This research has typically focused on *vertical* stratifications in education, of quantity of education and attainment differentials by students' background. With increasing access to university, relative *quality* of education, or the *horizontal* stratification within levels of education is an increasingly important driver of the intergenerational transmission of advantage (Gerber & Cheung, 2008). This chapter focuses on students' choices of field of study within university; building on the findings from chapter five showing that choice of subject was associated with family background.

Subject choices have strong implications for personal outcomes, including access to professional or higher paying occupations (Altonji, Kahn, & Speer, 2016; Walker & Zhu, 2011). They are also important for promoting an equitable society, which is compromised if students are stratified within education according to levels of advantage. Despite this, there remains limited research into the reasons for social background disparities in subject choices. In contrast, the mechanisms explaining gender segregation into subjects is a highly researched area, focussing primarily on the uneven distribution of personal traits that predict choices, including how much students enjoy subjects and their perceived ability in their chosen field (e.g. Sheldrake, Mujtaba, and Reiss, 2014; Eccles, 1983). This study extends the current literature by analysing relationships between students' attitudes, including their perception of ability and enjoyment of subjects, and subject choices at degree level.

Firstly, I consider whether differences in students' attitudes towards subjects can explain socio-economic gaps in subject choices at university. Whilst students' social background was associated with both subject studied at university, and their attitudes at age 13-14, differences in choices remained even for young people with similar attitudes. These differences also persisted when controlling for prior educational attainment and qualifications. The study goes further by examining whether students' attitudes are differentially associated with choices by students' family background. This could signal different drivers of choice for students from different social backgrounds, for example, whether students are less likely to choose subjects they enjoy or think they are good at, depending on family circumstances. I find that students whose parents had higher levels of education were more likely to choose STEM over arts and humanities as their enjoyment of STEM increased. Results are discussed with reference to the theoretical literature, and findings are contrasted with research into gender stratification into subjects.

6.2 Literature review

6.2.1 Field of study and social background

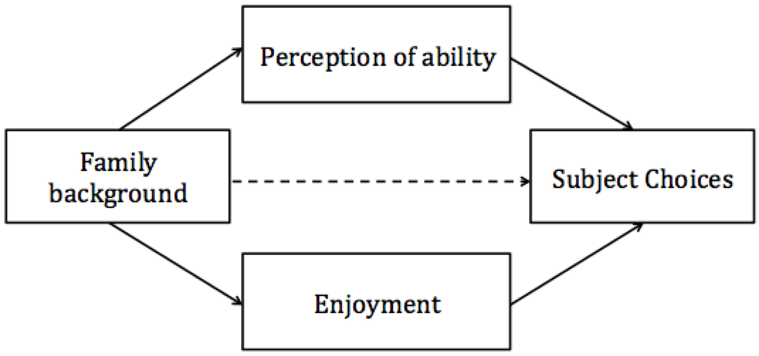
The literature on field of study in higher education has primarily considered a rather limited definition of subject choices, focusing on Science, Technology, Engineering and Maths (STEM) subjects over all other subjects. This reflects the large gender disparities in uptake of these subjects, and a strong policy agenda in increasing participation in STEM (e.g. HM Government, 2017). The higher education Statistics Authority (HESA) outlines key demographic characteristics associated with subject studied at university on a yearly basis, showing that socio-economic status (SES) disparities appear particularly large in Science, Engineering and Technology (CaSE, 2014).

These statistics do not, however, take into account attainment differences by students' background (e.g. The Royal Society, 2008). In response to arguments that these disparities occur because higher attaining students are both more likely to study STEM, and more likely to come from more advantaged families, a number of studies have explored the extent these disparities remain when accounting for differences in test scores. Van de Werfhorst, Sullivan, and Cheung (2003) analysed data from the 1958 National Childhood Development Study (NCDS), showing social class predicts participation in 'prestigious' subjects at university, i.e. medicine and law, even when attainment was taken into account. In contrast, Dilnot (2016) considered participation in subjects chosen at a much earlier age (16) which were most likely to facilitate entry to elite universities, finding that socio-economic trends in participation were largely explained by attainment and earlier choices. Henderson, Sullivan, Anders & Moulton (2016) found a similar social gradient in highly academically selective subjects and STEM subjects at age 14, which were again largely explained by attainment differences. Thus, the extent that associations persist after controlling for attainment may differ depending on the timing of choices. Focusing on science participation, Gorard & See (2009) exploited data from the Pupil-level Annual Schools Census (PLASC) and the National Pupil Database (NPD), exposing a strong association between SES and participation in all levels of post-compulsory science, and point out that no suitable explanation has been put forward to fully account for this disparity.

A key issue in identifying disparities by students' background lies in the measures used. Variously, studies focused either on social class, financial dis/advantage, parents' education, or some mixture of the three. In chapter five I showed that, compared with other background characteristics, parents' education plays the largest role in disparities in subject choice at university. Students whose parents are more highly educated are most likely to study arts and humanities and least likely to study social sciences, law and business.

There has been limited research into why these disparities in choices by social background occur. Whilst Gorard, See and Davies (2012) point to a need for more robust evidence in the question of how student attitudes and beliefs drive post-compulsory participation, there is a rich evidence base suggesting students' perception of ability and enjoyment could be key in explaining students' subject choices generally. In this chapter, I test whether differences in attitudes explain these disparities in subject choices. This hypothesised mechanism is illustrated in figure 6.1. Research in this area has typically focused on gender disparities in choice of STEM subjects, and has not considered either social background disparities or a broader range of possible choices.

Figure 6.1: Representation of the hypothesis that student's intrinsic motivations will explain differences in subject choices.



6.2.2 Perception of ability, enjoyment and subject choices

The relationship between perception of ability and STEM study has been explored in some depth; defined as the extent to which students rate their own ability positively either overall or in specific tasks. STEM subjects are perceived to be particularly difficult, and students perceive science and maths study to only be suitable for naturally 'brainy' students (DeWitt, Archer, & Osborne, 2013). Whilst there is indeed evidence science and maths are more difficult at A level when comparing relative difficulty of achieving high grades (Coe, Searle, Barmby, Jones, & Higgins, 2008), this additional barrier to study may put off many students who could otherwise enjoy STEM but are not confident in their academic ability. Whilst perception of ability is strongly related to actual attainment, it also independently predicts subject choices and aspirations. Results from the Programme for International Student Assessment (PISA) show that, across OECD countries, students' self-efficacy beliefs in mathematical problem solving at 15 is strongly associated with science career aspirations (Schulz, 2005). More recently, Sheldrake, Mujtaba, and Reiss (2014) show in a large longitudinal study that students' ratings of their ability in mathematics predicted both GCSE (age 16) attainment and aspirations for future study. Students' self-beliefs can also go some way to explaining gender disparities in subject choices. Girls' relatively low academic self-concept, compared to boys, can go a large way to explaining the underrepresentation of young women in STEM (e.g. Lyons and Quinn 2010). This was particularly pronounced in 'harder' physical sciences, in which the largest gender disparities in participation are observed. It is unclear whether perception of ability can explain disparities in choices along other student characteristics, which this chapter aims to address by looking directly at disparities by social background.

Along with perception of ability, Sheldrake et al. (2014) also found intrinsic motivation to be key in aspirations for future mathematics study. Intrinsic motivation, or students' inclination to study subjects based on personal reward and enjoyment, is an important factor in academic decisions. Whilst it seems that overall students do enjoy studying mathematics at the start of secondary school, there is considerable variation in preferences, and enjoyment appeared to be declining in line with future study aspirations from 2003-2007 (National Audit Office, 2010). Where this relates to students' family background is less clear.

The study of intrinsic motivations has strong roots in psychological literature. Eccles aimed to explain gender differences in uptake of science and mathematics by modelling psychological characteristics of students, and their subsequent choices (Eccles, 1983). There has been extensive research into associations between subjective task-value and subject choice, finding consistently that task-value can go some way to explaining gender gaps (e.g. Eccles, 2011; Eccles & Wigfield, 2002). In contrast, as part of another longitudinal study into student aspirations with a focus on STEM, ASPIRES, DeWitt, Osborne, et al. (2013) show enjoyment of science and mathematics do not necessarily predict participation. However, little work focuses on SES, and research in this area was generally undertaken with students from more advantaged backgrounds.

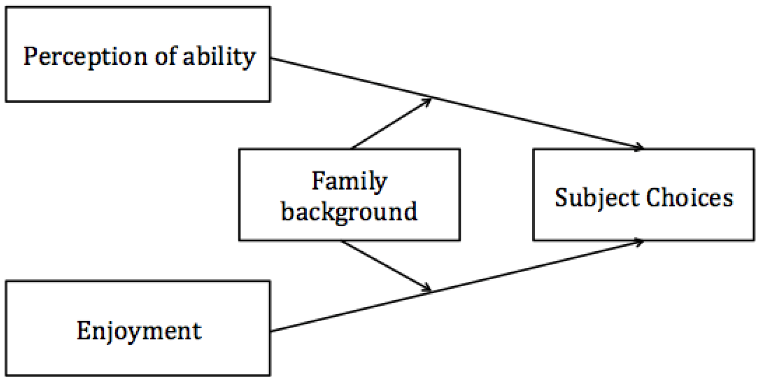
6.2.3 Different drivers of choice?

Attitudes may also be differentially related to choices depending on students' family circumstances. This possible moderation effect is illustrated in figure 6.2. Cultural reproduction theory (Bourdieu, 1984) and the theory of relative risk aversion (Breen & Goldthorpe, 1997), offer some insight into differing processes underlying choice depending on students' background. Cultural reproduction theories focus on cultural capital held by more advantaged families, including education, cultural knowledge and participation, and manner of speech and presentation. It appears parents' education specifically drives disparities in subject choices, as opposed to financial resources or social class (Codiroli Mcmaster, 2017), and parents who have been to university may have more knowledge of the range of options available for students within university, and the career opportunities those options may lead to. They may also be more likely to encourage students in their interests through involvement in their education (Sacker, Schoon, & Bartley, 2002; Sui-Chu & Willms, 1996), and through promoting after school activities that match their preferences (Lareau, 2000).

Researchers working on the ASPIRES project suggest that another form of capital, students' science capital (the extent to which their families have knowledge of STEM, work in STEM careers themselves and encourage STEM participation), may account for participation disparities. Science capital is strongly associated with other forms of capital and students' relative level of advantage (Archer et al., 2012). Students with more science capital are more likely to be knowledgeable about the range career options after studying STEM, and to realise that skills learned from STEM degrees can be transferable to many different sectors and roles. This echoes Akerlof (1997), who argued parents pass on knowledge of university systems. In respect to subject choices, parents appear to pass on knowledge of the value of studying particular subjects, and the relative advantages they may confer.

According to the model of Relative Risk Aversion, people aspire to achieve social standing that is at least as good as their parents (Breen & Yaish, 2006), leading to lower educational aspirations if parents aren't well educated themselves. Assuming students want to avoid downward social mobility, this may lead students from more advantaged backgrounds to aspire to more prestigious subjects, and to be more concerned with economic return to study over subjects they enjoy. Recent UK research suggests that students from higher income families are indeed more concerned with economic returns of university choices (Davies, Mangan, Hughes, & Slack, 2013). For this to translate into to more advantaged students choosing higher return subjects rests on the assumption that students have accurate understandings of returns to education (Botelho & Pinto, 2004; Manski, 1993).

Figure 6.2: Representation of the hypothesis that student's intrinsic motivations have different relationships with subject choices depending on their backgrounds (moderation effect)



In contrast, considering the additional barriers students from less advantaged background face in the labour market (Crawford & Greaves, 2015), more advantaged students may see university as a chance to study something they're interested in 'for learning's sake,' and to be more concerned with intrinsic rewards university study will bring over extrinsic rewards. In his 1974 book, Boudon outlined the differences between primary and secondary effects in education. Primary effects refer to attainment in school, which may influence subjects students *can* study, whilst secondary effects refer to choices made by students based on values and preferences passed down by parents (Boudon, 1974; Girard and Bastide, 1963). Boudon argues that secondary effects of social background on education arise from the fact that there are different benefits, and costs, to remaining in education depending on family resources (1974). Whilst it is likely the majority of students are somewhat concerned about job security and salaries upon graduation, this may be a more salient concern for students from less educated families, who have less of a 'safety net' provided by parents and family. They may not have access to professional networks, knowledge and/or financial capital to help enter more stable professions (particularly arts and humanities focused jobs). In other words, there is a higher cost and lower benefit to obtaining a degree for less advantage students, who may therefore be *less* concerned about choosing subjects they enjoy, and more concerned about selecting a subject with the highest possible return. This echoes some previous studies suggesting students from higher SES backgrounds are more concerned with intrinsic, rather than extrinsic, rewards of higher education (Kohn & Schooler, 1983; Mortimer, Lorence, & Kumka, 1986).

In summary, previous literature tells us that students' background is associated with subject choices, with students from more privileged backgrounds appearing to choose subjects that confer higher economic rewards and potentially entrenching their privilege. Another strand of research suggests student' attitudes and preferences strongly predict choices, and can go a large way to explaining choice disparities by gender. This chapter brings together these distinct literatures to further understand the reasons for disparities in subject choices by family background. I test the hypothesis that differences in students' choices are driven by differences in students' personal attributes, specifically ratings of their own abilities and enjoyment in these subjects. I do this by looking at students' choices of three groups of subjects: STEM; Social sciences, Law and Business (SLB); and arts and humanities. SLB subjects are distinguished from arts and humanities subjects because they offer very different occupational returns upon graduation and different students choose these subjects (Walker & Zhu, 2013; Codiroli Mcmaster, 2017).

6.3 Method

6.3.1 Participants

This chapter uses data from Next Steps. Full details of the dataset are given in chapters four and five. Of the 8,682 participants in wave seven, 3,894 were studying for a degree at university, and of these 3,884 gave valid responses for subject studied, and 3,878 also gave valid responses for ethnicity and gender. Thus, the final analytical sample was 3,878. As in the previous chapter, combined longitudinal and cross-sectional analytical weights were used throughout analysis (unless indicated otherwise).

The longitudinal nature of the data allowed me to compare students' characteristics and attitudes measured at 13 - 14 (in the first wave of data collection), with choices at 18-19, eliminating the possibility that subject studied would influence reporting of characteristics or attitudes. For example, students who were studying STEM, or were in the process of applying to study STEM at university, may report enjoying science and maths because they were more actively engaged with the subjects.

6.3.2 Analytical strategy

This chapter aims to address the following two research questions:

1. Do students' enjoyment and perception of ability explain observed disparities in subject choice by student's background?
2. Do these associations differ by students' parents' education level?

I first present descriptive statistics outlining the extent of differences in subject choices and attitudes towards STEM and English by students' parents' education level, and students' relative attitudes by subject choices. Examining the raw relationships between attitudes, family background and subject choices is an important first step in informing my first research question. I go on to present a series of tables outlining the proportions of students studying each subject group at different levels of perception of ability and enjoyment in STEM and English, split by parents' education level. This will go some way to answering my second research question: do associations between student's attitudes and subject choices differ depending on their social background?

Raw comparisons of proportions of students studying each subject can give some indication of disparities in participation, however they do not give the full picture since student characteristics are highly correlated with one another, and also with prior achievement in school. For example, students' ethnicity and social background are both highly correlated, and associated with choices in different ways. In consideration of this, and to more fully address my research questions, I use multinomial logistic regression methods. The regression models are built up in four stages with increasing levels of complexity, described in detail in the results section.

6.3.3 Key variables and descriptive statistics

Subject choice

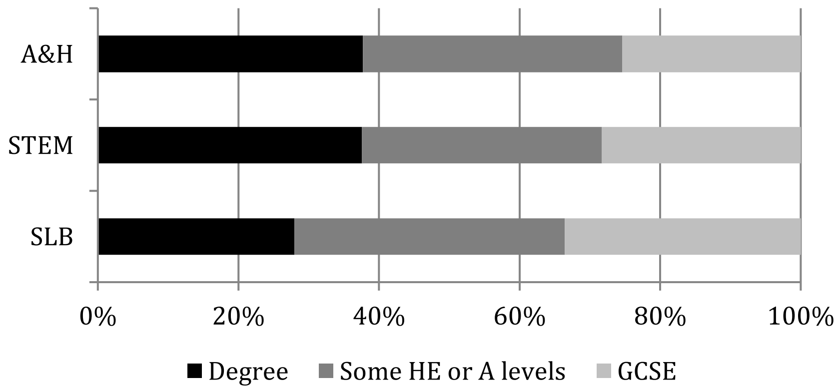
The approach used to classify subjects is the same as that outlined in chapter five.

Student characteristics

The focal measure of family background in this study is parents' highest qualification; following findings from chapter five that parents' education has the strongest association with subject choices, when compared with parent's social class and financial resources. The qualification of the parent with the highest education level (or only parent) was used in analysis, and students were split into three groups: those whose parents had a degree or higher qualification (35%), those whose parents had A levels, some higher education or equivalent (36%), and those whose parents were educated to GCSE level or below (29%). This was taken from wave one interviews with parents, at the same time as students' reports of enjoyment and perception of ability in subjects. Students' ethnicity (white, mixed ethnicity, Indian, Pakistani, Bangladeshi, black African, black Caribbean, or other ethnicity) and gender were also included in all analysis.

Figure 6.3 shows the raw relationship between students' parent's education level and subject choices; students whose parents are better educated are most likely to choose either STEM or arts and humanities (A&H) subjects, and least likely to choose SLB subjects.

Figure 6.3: Parents' highest qualification by subject chosen



Enjoyment and perception of ability

Enjoyment of STEM was defined by combining two variables: ‘How much do you like or dislike this subject: maths,’ and ‘How much do you like or dislike this subject: science.’ For enjoyment of English, students were asked ‘How much do you like or dislike this subject: English.’ Ratings were on a 1-4 Likert scale, with 1 indicating ‘like a lot’ and 4 indicating ‘don’t like at all.’ Scores were reversed, so a high score indicates high enjoyment of each subject. Attitudes towards maths and science were combined to reflect the fact that choice of studying maths and science were combined in the outcome measure. Perception of ability in STEM was defined by combining scores for questions ‘How good or bad [are you] at this subject: maths?’ and ‘How good or bad [are you] at this subject: science?’ For English, students were asked ‘How good or bad [are you] at this subject: English?’ These ratings were also on a 1-4 likert scale, with 1 indicating ‘very good’, and 4 indicating ‘very bad’. For the final variable, high scores indicated high perception of ability. These traits were measured in the first waves of data collection, when students were 13 - 14 years old.

Figures 6.4 & 6.5 show how students’ enjoyment of, and perception of ability in STEM and English differ by parents’ education level and by the subjects they study at university. Students whose parents are highly educated are most likely to say they are good at STEM and English, and (to a lesser extent) to say they enjoy science, maths and English. Students studying STEM subjects at university were most likely to say they were good at, and enjoyed, science and maths at 13-14, whilst those who choose arts and humanities were most likely to have said they were good at, and enjoyed English at school.

Figure 6.4: Student standardised enjoyment and perception of ability in STEM and English at age 13-14 by parent education level

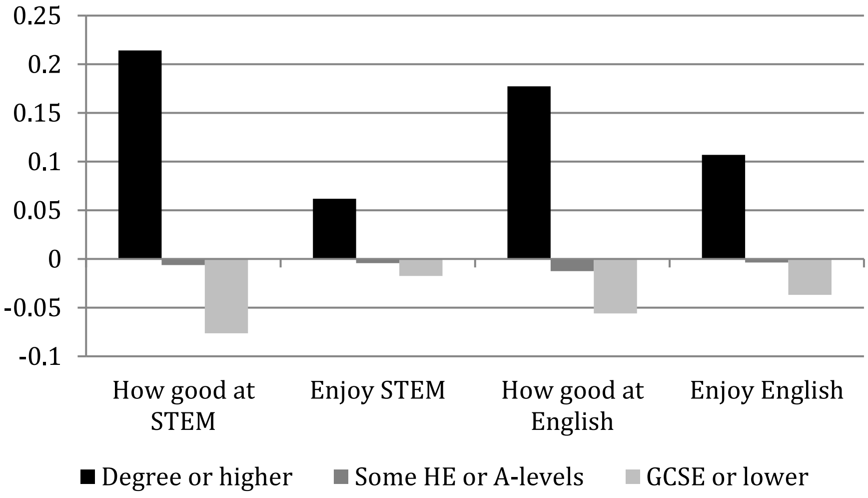
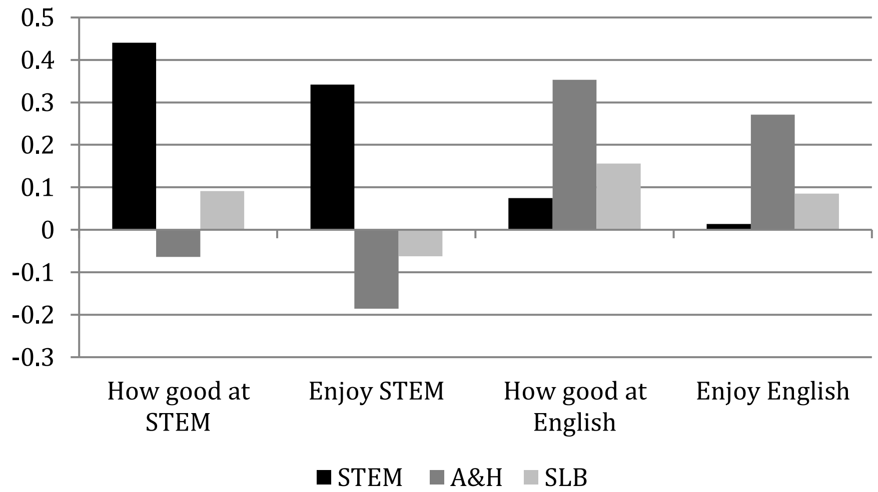


Figure 6.5: Student standardised enjoyment and perception of ability in STEM and English at age 13-14 by subject studied in university



Prior attainment and qualification type

Students' prior attainment measured at KS2 and KS4 (GCSEs) are included as controls in the analysis. KS2 point scores in maths, science and English are included separately to acknowledge expected differing associations between achievement and choice across the three subjects. Due to data restrictions, GCSE scores could not be included as separate subjects, so capped overall scores are included.

Attainment at A level (or equivalent) was not included in consideration that difficulty of subjects differs, and students would have taken very different profiles of subjects (with different levels of difficultly) depending on the subject group they aspired to study at university. Type of qualification was included, however, coded as A level or 'other qualification.' 84% of young people attending university had studied A levels. The remaining qualification types were combined to retain sample sizes, but primarily include more vocationally oriented qualifications (e.g. Business Technology Education Council (BTEC) qualifications).

All continuous measures were standardised to mean 0 and standard deviation 1.

6.4 Interactions between social background and attitudes

Figures 6.6-6.9 illustrate how associations between attitudes and subject choices differ by parental education, with each line representing students whose parents have a different level of education. Students are split into three equal sized groups (low, medium or high) according to their perception of ability or enjoyment in STEM and English in relation to their peers. The vertical axis represents the proportions of students studying the select subject groups. Where lines diverge differences in subject choices by social background are observed, and where lines are not parallel interactions between social background and attitudes are observed. Overall results suggest that students whose parents have a degree are more likely to be driven by how good they think they are and how much they enjoy STEM or English in making subject choices, compared to students whose parents have lower levels of education.

Figures 6.6 and 6.7 compare the proportions of students studying each subject group by their perception of ability and enjoyment of STEM. For choice of arts and humanities subjects, students of different social backgrounds who think they are good at, or enjoy STEM are more alike in their choices (in this case with lower proportions choosing arts and humanities), and differences in choices appear larger for students who do not think they are good at, or like STEM. For choice of STEM subjects a similar interaction is observed in the opposite direction. As students' perception of ability and enjoyment of STEM increases the social gradient in choices increases. There does not appear to be a consistent interaction concerning SLB subjects and perception of ability, however as enjoyment of STEM subjects increases, students whose parents have a degree or higher are increasingly less likely to study SLB compared to students whose parents have lower levels of education.

Figures 6.8 and 6.9 show that similar interactions are also observed for perception of ability and enjoyment of English. Overall, the associations between attitudes and choice of arts and humanities or STEM subjects are stronger for students whose parents have higher levels of education. These patterns are again not observed for choice of SLB subjects.

In the introduction I discussed possible directions of interactions suggested by the theory of relative risk aversion. According to the theory, students from more advantaged backgrounds would be more likely to choose subjects that had higher occupational returns in aiming to avoid downward mobility, and less likely to choose subjects based on intrinsic motivations. In contrast, students from less advantaged backgrounds are already achieving upwards mobility simply by attending university. The data suggests students whose parents are more educated are most likely to choose subjects for intrinsic reasons. This is contrary to what would be expected if the theory of relative risk aversion were applicable to subject choices.

As the main driver of disparities is parents' education level specifically, rather than social class or family income, an alterative interpretation is that parents' education more directly affects the strength of associations. The literature suggests more educated parents are better able to foster students' interests and perceived strengths, and push them in the direction of subjects that suit their individual preferences (Sacker, Schoon, & Bartley, 2002; Sui-Chu & Willms, 1996; Lareau, 2000). This explanation would account for the fact that the results run in the opposite direction to what would be expected given the theory of Relative Risk Aversion.

Results are, however, in line with Boudon's (1974) arguments about the different costs and benefits to higher education, which may then lead to different drivers behind these decisions. Whilst there is more chance of occupational success upon graduation for students who study STEM or SLB, more advantaged students may also expect a level of success from studying arts and humanities, and their less advantaged peers may face more barriers upon graduation in these particular subjects. For example, they will have more access to well-educated networks that can offer advice and guidance in applications and work experience. Their parents will also be more able to support them financially through periods of worklessness or unpaid internships. Further, their increased cultural capital may help them indirectly, and be particularly useful when applying for and attending interviews for jobs in arts and humanities. This could thus explain why students from more advantaged backgrounds are more inclined to study subjects for 'enjoyments sake,' and worry less about employability upon graduation.

These tables show a consistent picture of differences in associations between students' attitudes and choices by their background. In the next stage of analysis I go on to test whether relationships remain when controlling for other factors, including prior attainment.

Figure 6.6: Proportion of students studying each group of subjects by perception of ability in STEM, split by parents' education

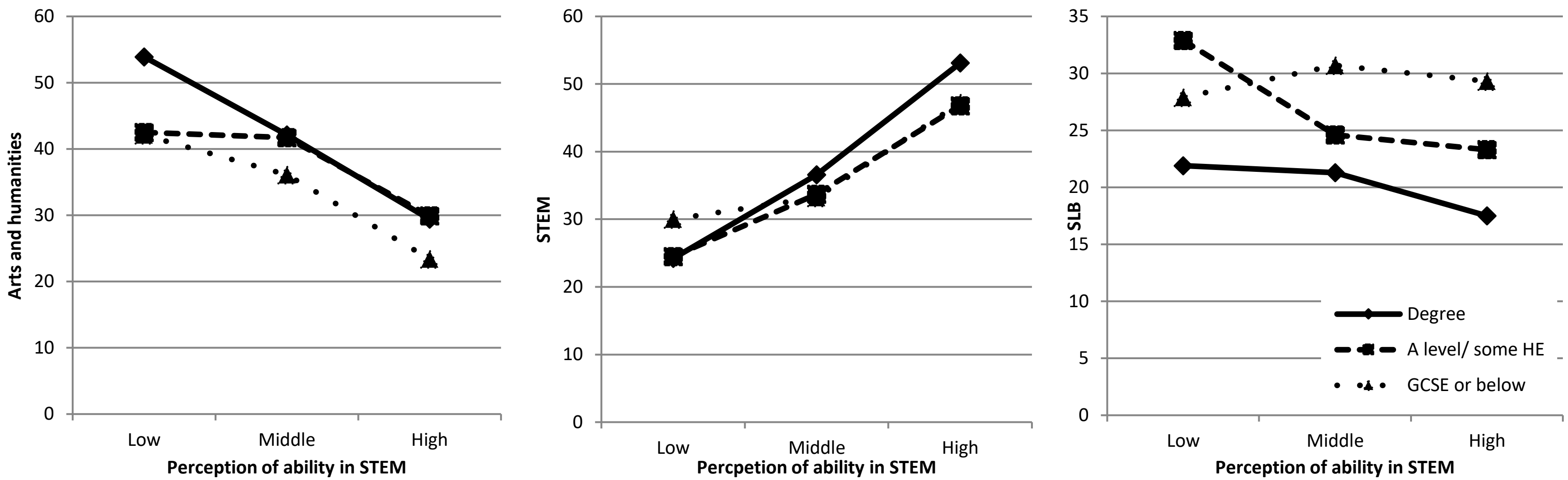


Figure 6.7: Proportion of students studying each group of subjects by enjoyment of STEM, split by parents' education

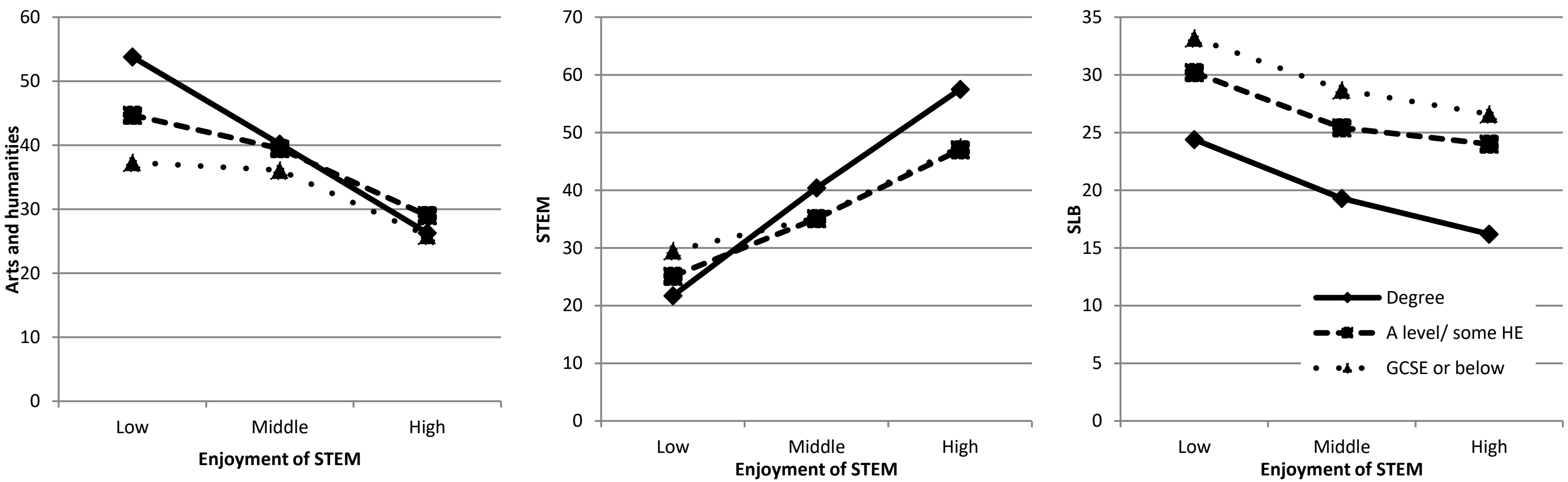


Figure 6.8: Proportion of students studying each group of subjects by perception of ability in English, split by parents' education

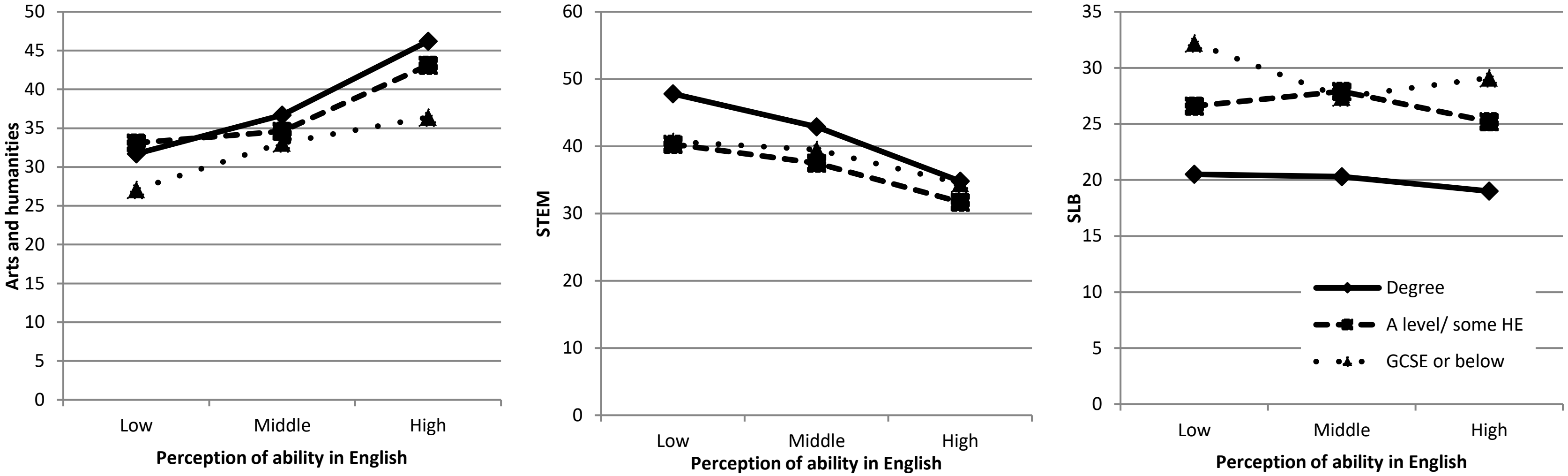
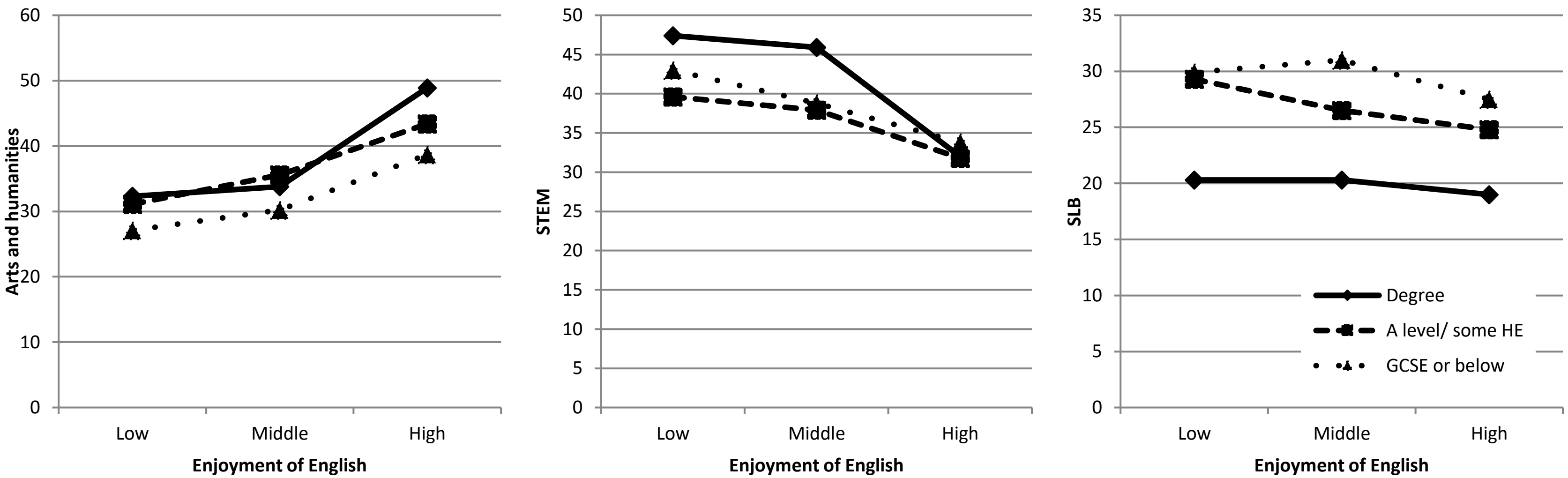


Figure 6.9: Proportion of students studying each group of subjects by enjoyment of English, split by parents' education



6.5 Regression results

The analytical strategy employed in this chapter mirrors that used in chapter five; I built up a series of multinomial regression models with increasing complexity. The first multinomial logistic regression model aimed to test the 'raw' association between parents' education level and subject choices, with ethnicity, gender and other family background characteristics controlled. For simplicity, and in contrast to the approach used in chapter five, I include a single measure of parental education rather than including parents separately. Despite this, results are substantively consistent across chapters. Relative risk ratios from table 6.1 illustrate that students whose parents have higher levels of education are more likely to choose STEM over SLB subjects at university. Students' whose parents have lower education levels were around 46% more likely to choose SLB subjects than STEM subjects, compared with students whose parents had a degree or higher qualification.

The second and third models control for academic attainment and attitudes, which change associations substantially. The introduction of these variables aims to answer my first research question: Do students' enjoyment and perception of ability explain observed disparities in subject choices by student's background? For students with similar academic attainment, enjoyment, and perception of ability, only students whose parents have intermediate levels of education remain more likely to choose SLB over STEM, compared with students whose parents have high levels of education. The difference in choices between STEM and SLB subjects for students with the lowest and highest levels of education is no longer statistically significant. The relationship between social background and choice of arts and humanities over STEM subjects, however, becomes statistically significant when controlling for GCSE scores. This suggests that academic attainment, particularly attainment at 16, is acting as a stronger push factor to studying STEM subjects instead of arts and humanities for less advantaged students.

Students' perception of ability and enjoyment of subjects did indeed predict university choices over and above their relationship with prior attainment. Attitudes are standardised so relative risk ratios represent change in propensity to study arts and humanities or SLB, over STEM, with 1 standard deviation increase in the corresponding indicator. Students studying arts and humanities subjects rated themselves as less able in maths and science, and more able in English, at 13/14 than their peers who choose STEM subjects. They also said that they enjoyed maths and science less, and although they enjoyed English more this relationship was only significant at the 10% level. Students who chose SLB subjects also thought they were less able in maths and science than students who choose to study STEM, but they enjoyed these subjects more and enjoyed studying English less.

The profile of students studying arts and humanities over STEM subjects indicated by the models is not surprising. It includes students who, relative to their peers, think they are good at English, and are both less able at STEM as well as not enjoying the subjects as much. In contrast, the profile of students studying SLB subjects over STEM offers some interesting insights into student choices. Compared to their peers, they seem to enjoy maths and science, but do not see themselves as good at the subjects, and come from families with intermediate levels of education. It is perhaps the case that these students are put off by the perception that maths and sciences are particularly difficult subjects, thus choose subjects that may have some STEM content but are seen as more accessible regardless of ability. That these relationships are observed for students with similar academic ability indicates that students' perception of their ability, over and above that informed by their actual test scores, is driving their choices.

In model three I additionally include an indicator for whether students studied A levels, or alternative examinations, pre university entry. Whilst the majority of university students in the sample studied A levels, in the UK education system students who are aiming to study more vocationally focused courses at university may study BTEC examinations. Entry into these alterative examinations is associated with prior attainment and social background. Thus, we may expect students who are channelled into these more vocational routes to not only study different subjects than their more 'academic' peers, but to also be less likely to choose subjects based on intrinsic values. As Boudon (1974) argued, these alternative branching points available in the UK education system are likely to increase class disparities in educational choices, as more advantaged young people (and their parents) are better able to use these choices to their advantage. Students who studied A levels were more likely to study SLB than those who achieved other qualifications pre-entry, however, inclusion of this variable did not affect the coefficients on the relationship between parental background and subject choice. I also considered another branching point, entry into Russell Group universities, which include primarily research focused institutions. Similarly, the inclusion of this variable did not substantively impact other associations, but was associated with subject choices (people attending a Russell Group university were more likely to study STEM over both arts and humanities and SLB). The fact that the relationship between subject choice and family background remains significant when accounting for these different qualification and university types suggests that results are not solely driven by the streaming of students into vocationally oriented pathways.

Finally, my second research question was whether these associations differ by students' parents' education level. Whilst this has been addressed in the descriptive statistics, in figures 6.6-6.9, the fourth regression models were run to test whether interactions persisted, and were statistically significant, when controlling for other student characteristics and attainment. The interaction between students' enjoyment of STEM subjects and their parent's education in choice to study arts and humanities over STEM was statistically significant. As enjoyment of STEM increased by a standard deviation, students whose parents had GCSEs or lower were around 50% more likely to choose art and humanities over STEM than students whose parents had a degree or higher. This interaction was robust to associations with prior academic attainment, qualifications and university type. Other interactions, including between students enjoyment of English, and perception of ability in STEM or English, were not statistically significant when accounting for other background characteristics, including attainment and university attended.

Table 6.1[14]: Multinomial logistic regressions showing students odds of studying arts and humanities, or SLB, over STEM subjects at university. Relative risk ratios are shown with standard errors in parenthesis.

Subject choice (reference: STEM)	Arts and humanities				SLB			
	Raw	Attitudes and KS2 attainment	All prior attainment and qualifications	Interactions	Raw	Attitudes and KS2 attainment	All prior attainment and qualifications	Interactions
Parents Education								
Reference: Degree								
HE or A levels	1.034	0.958	0.872	0.881	1.511***	1.439***	1.294**	1.254
	(0.109)	(0.107)	(0.101)	(0.113)	(0.188)	(0.182)	(0.168)	(0.176)
GCSE or lower	0.965	0.822	0.726**	0.721**	1.461***	1.310*	1.141	1.074
	(0.114)	(0.103)	(0.0947)	(0.103)	(0.193)	(0.182)	(0.162)	(0.166)
Attitudes towards subjects								
How good at STEM		0.716***	0.737***	0.705***		0.843**	0.869**	0.911
		(0.0455)	(0.0474)	(0.0795)		(0.0577)	(0.0601)	(0.115)
Enjoy STEM		0.740***	0.751***	0.642***		0.693***	0.706***	0.617***
		(0.0432)	(0.0445)	(0.0676)		(0.0444)	(0.0456)	(0.0755)
How good at English		1.398***	1.435***	1.553***		1.131*	1.161**	1.200
		(0.0867)	(0.0896)	(0.172)		(0.0722)	(0.0742)	(0.146)
Enjoy English		1.117*	1.127**	1.172		1.014	1.020	1.068
		(0.0659)	(0.0676)	(0.135)		(0.0589)	(0.0599)	(0.122)
Parents education x STEM attitudes								
Degree x Good at STEM								
HE or A levels x Good at STEM				1.199				0.795
				(0.182)				(0.143)
GCSE or lower x Good at STEM				0.922				1.137
				(0.156)				(0.191)
Degree x Enjoy STEM								
HE or A levels x Enjoy STEM				1.140				1.311
				(0.163)				(0.220)
GCSE or lower x Enjoy STEM				1.495***				1.141
				(0.229)				(0.185)
Parents education x English attitudes								
Degree x Good at English								
HE or A levels x Good at English				0.874				1.073
				(0.132)				(0.178)

[14] Other family background characteristics were included in initial analysis to ensure that parents' education was the main driver of choices, and that coefficients did not change substantially if they were added as controls. These included NS-SEC occupational social class (highest of both parents), housing tenure, and how well the family reported managing on finances. None were independently associated with subject choices, nor did they substantively impact results. Thus, the more parsimonious regression models are presented in this article.

GCSE or lower x Good at English				0.894				0.825
				(0.144)				(0.138)
Degree x Enjoy English								
HE or A levels x Enjoy English				0.921				0.853
				(0.138)				(0.130)
GCSE or lower x Enjoy English				0.964				1.035
				(0.153)				(0.161)
Controls								
Female	1.585***	1.082	1.139	1.135	1.196*	0.927	0.988	0.996
	(0.139)	(0.103)	(0.111)	(0.111)	(0.114)	(0.0972)	(0.106)	(0.107)
Ethnicity								
Reference: White								
Mixed	1.014	1.085	1.099	1.124	1.303	1.345	1.366	1.361
	(0.247)	(0.280)	(0.300)	(0.309)	(0.356)	(0.379)	(0.389)	(0.390)
Indian	0.438***	0.462***	0.463***	0.461***	1.535***	1.594***	1.596***	1.628***
	(0.0759)	(0.0876)	(0.0883)	(0.0885)	(0.219)	(0.244)	(0.250)	(0.253)
Pakistani	0.245***	0.247***	0.244***	0.232***	1.457*	1.545**	1.519**	1.518**
	(0.0617)	(0.0598)	(0.0601)	(0.0570)	(0.283)	(0.297)	(0.295)	(0.294)
Bangladeshi	0.492***	0.485***	0.486***	0.473***	1.693***	1.760***	1.747***	1.728***
	(0.121)	(0.120)	(0.124)	(0.118)	(0.328)	(0.374)	(0.363)	(0.350)
Black Caribbean	0.847	0.642	0.604	0.609	1.339	1.200	1.125	1.178
	(0.304)	(0.280)	(0.262)	(0.263)	(0.475)	(0.455)	(0.420)	(0.432)
Black African	0.774	0.664	0.672	0.651	2.151***	2.117***	2.164***	2.100***
	(0.196)	(0.176)	(0.183)	(0.176)	(0.487)	(0.511)	(0.542)	(0.524)
Other	0.383***	0.405***	0.423***	0.413***	0.844	0.938	1.001	1.000
	(0.117)	(0.131)	(0.138)	(0.139)	(0.226)	(0.256)	(0.277)	(0.278)
Independent school	0.896	0.671**	0.752*	0.764	1.058	0.909	1.016	1.015
	(0.139)	(0.111)	(0.130)	(0.135)	(0.192)	(0.171)	(0.197)	(0.198)
Academic attainment								
GCSE (overall capped score)			0.739**	0.737**			0.667***	0.668***
			(0.110)	(0.110)			(0.0996)	(0.0998)
KS2 Maths score		0.735***	0.789**	0.784**		0.990	1.075	1.081
		(0.0684)	(0.0754)	(0.0753)		(0.100)	(0.114)	(0.115)
KS2 Science score		0.946	0.973	0.973		0.797**	0.827*	0.820*
		(0.100)	(0.105)	(0.106)		(0.0887)	(0.0923)	(0.0918)
KS2 English score		1.027	1.123	1.146		1.031	1.143	1.145
		(0.106)	(0.121)	(0.123)		(0.109)	(0.129)	(0.129)
Studied A-levels			1.235	1.235			1.400**	1.382**
			(0.183)	(0.184)			(0.209)	(0.207)
Attending a Russell Group university			0.592***	0.593***			0.552***	0.552***
			(0.0694)	(0.0706)			(0.0713)	(0.0719)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

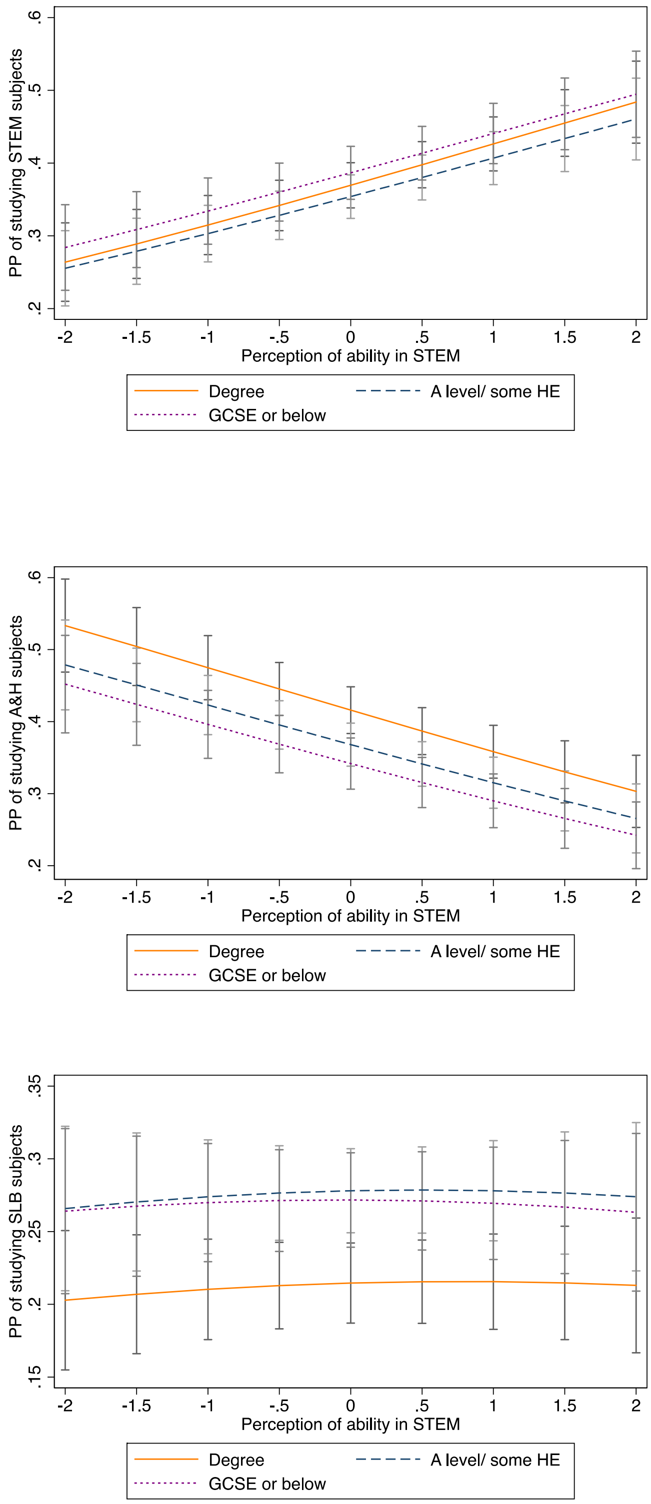
6.6 Interactions between social background and attitudes: full model

In the final model of the regression, interaction terms between students' parents' level of education, and attitudes towards subjects were included. When controlling for other associated characteristics, only interactions between parents' education and enjoyment of STEM subjects were statistically significant. Students whose parents had higher levels of education were more likely to choose STEM subjects over arts and humanities as their enjoyment of STEM subjects increased, than students whose parents had low levels of education. The interaction between parents' education and enjoyment of STEM subjects is also significant, however only at the 10% level.

The disparity in choice of STEM over SLB subjects increased with students' enjoyment of STEM, such that students whose parents had a degree were increasingly more likely to study STEM compared with students whose parents had A levels or some higher education as their enjoyment of STEM increased. This suggests that there is a difference in the gradient of the slope in the relationship between students' enjoyment of STEM and subject choice, with a larger difference in choices by social background for students who enjoy STEM, and a smaller social difference for students who do not enjoy STEM. Thus in line with descriptive analysis, students whose parents are more educated appear to be more likely to choose subjects based on intrinsic motivations, particularly on what they enjoy studying

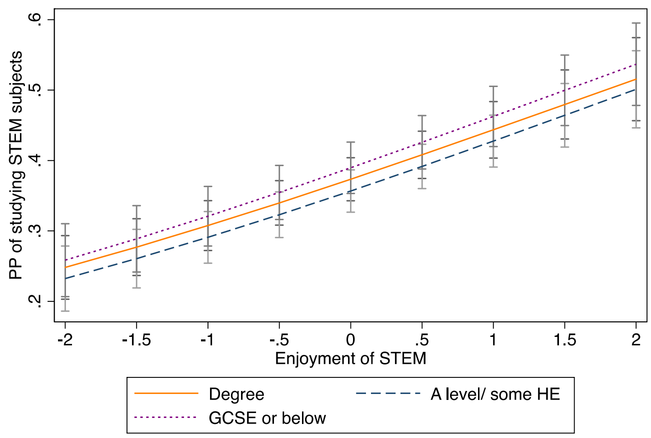
Figures 6.10 & 6.11 illustrate the predicted probabilities of students choosing STEM, arts and humanities, or SLB subjects by students' perception of ability and enjoyment of STEM subjects (measured as standards deviation differences from the mean). They give an alternative view to the relative risk ratios discussed above, as they predict probability of studying each group of subjects individually, rather than in comparison with one another. Relationships are estimated at each level of parental education to assess the differences in slopes of the lines. If slopes diverge, differences in the social gradient in subject choice by students' enjoyment of STEM are observed.

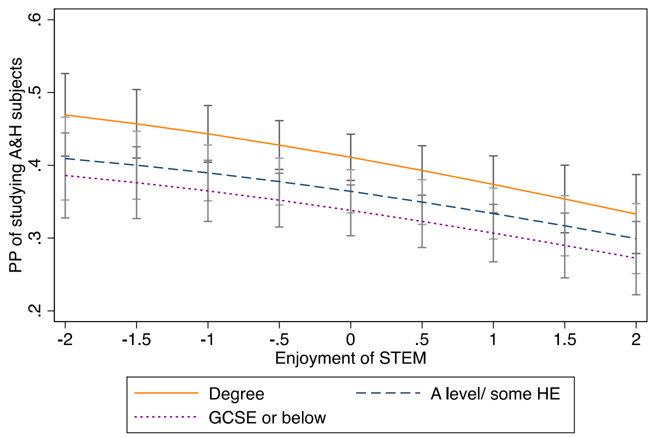
Figure 6.10: Predicted probability of studying each group of subjects by perception of ability in STEM and parents' education level



The relationships between students' social background and subject choices by perception of ability in STEM are shown in figure 6.10. Whilst there is little difference in propensity to study STEM subjects by parent's education, students whose parents have a degree are more likely to study arts and humanities and less likely to study SLB subjects than students whose parents have lower qualifications. Students who think they are good at STEM subjects in school are more likely to choose STEM and less likely to choose arts and humanities. There is little association between choice of SLB subjects and perception of ability; students appear equally likely to study SLB whether they think they are good at STEM or not.

Figure 6.11: Predicted probability of studying each group of subjects by enjoyment of STEM and parents' education level





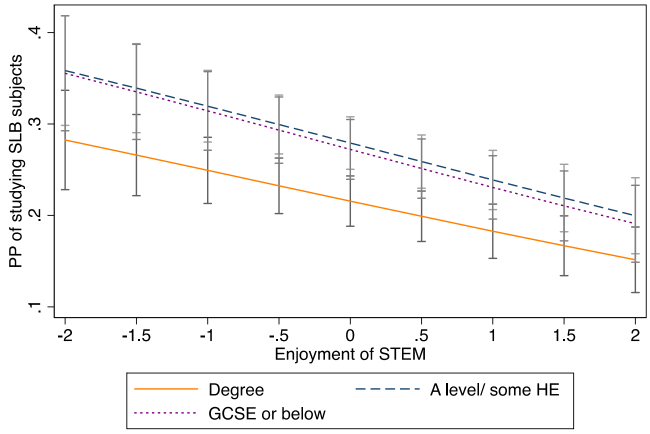


Figure 6.11 shows that students whose parents have a degree are most likely to be studying STEM subjects, and students whose parents have intermediate levels of education least likely. A similar relationship is also seen for arts and humanities subjects, and students with highly educated parents are least likely to study SLB subjects. Whilst the gradient of the three lines appears similar for all students studying STEM subjects regardless of parent's education level, they differ for propensity to study arts and humanities or SLB subjects. For arts and humanities subjects the gradient is steeper, suggesting that enjoyment of STEM subjects has a stronger negative association with choices for students whose parents have a degree, than for students whose parents have lower levels of education. Whilst advantaged students remain more likely to study arts and humanities at university, for those who enjoy STEM the confidence intervals overlap, suggesting differences are no longer significant. For choice of SLB subjects, the social disparities are highest for students who do not enjoy STEM, suggesting that for students who don't like STEM, students whose parents have higher levels of education are particularly less likely to study SLB than students whose parents have lower levels of education. This is perhaps reflected in the fact that studying arts and humanities is negatively associated with social background, and students who dislike maths and science but do not have highly educated parents are more likely to choose SLB subjects instead.

6.7 Discussion

This chapter explored the mechanisms of horizontal stratification in the English education system of students into different subject areas. Historically, academics have focused on vertical stratifications in education by social background, or gender differences in subject specialisations; however, more recent research has shed light on differences in subjects studied by social background (Anders, Henderson, Moulton, & Sullivan, 2018; Codiroli Mcmaster, 2017; Dilnot, 2016; Van de Werfhorst et al, 2003). In chapter five, I found that social background was associated with subject choices for a cohort of young people born between 1989 and 1990, however, the reasons for these differences in choices remained unclear. This chapter set out to explore whether a common explanation for gender differences in choices, differences in attitudes towards subjects, could also be applied to differences in choices by social background.

Alongside work considering how student characteristics influence subject choices, a distinct area of research identified a selection of personal attitudes that predict choices. Students' subjective task-value, the extent to which students *want* to study a subject, and beliefs about their own ability were identified as important drivers of choice. Subjective task value can be split into four sub-components, with 'intrinsic value,' or the extent that students enjoy a subject, being a crucial factor (Eccles, 1983). These studies were typically based on relatively advantaged students, and the extent to which these attitudes and associations differ based on students' characteristics, with exception of gender, had not yet been comprehensively explored.

This chapter adds to the literature on the psychological mechanisms informing choices by analysing a large, representative cohort of university students across England from a range of social backgrounds. I examined the relationship between students' personal and background characteristics in determining subject choices at university, and whether students make subject choices for the same reasons regardless of background. The study replicated prior work by showing differences in the subjects students chose to study according to their parents' education level. Students whose parents had higher levels of education were both more likely to choose arts and humanities, and less likely to choose social sciences, law or business, compared to students whose parents had lower levels of education. The study also confirmed findings from the psychological literature, showing that students from a range of social backgrounds were most likely to choose subjects they thought they were good at and enjoyed.

A unique contribution of this research is that it shows that ratings of enjoyment and perception of ability influenced university choices over and above actual attainment and a range of other student characteristics. Furthermore, the uneven distribution of students' enjoyment of, and confidence in subjects by background was considered a potential driver of disparities in subject choices. Descriptive statistics suggested students whose parents had lower levels of education were less likely to enjoy science and maths and to rate themselves as 'good' at these subjects. However, disparities in uptake remained when controlling for attitudes. Even when students enjoyed STEM, and thought themselves equally capable, students whose parents were more educated remained more likely to study arts and humanities over STEM, and to study STEM over SLB subjects at university. In contrast, gender differences in subject studied at university were explained entirely by differences in attitudes. More research is required to understand fully why the processes driving gender and social background disparities in choices differ so widely.

Whilst initial results confirmed positive associations between attitudes and subject choices, further analysis sought to understand whether all students, regardless of background, were equally likely to make choices based on their personal preferences and beliefs about their abilities. The chapter identifies differences in processes influencing choices by students' background. As students' perception of ability and enjoyment of STEM increases, the social gradient in choices increases. To further understand this, it is important to explore the specific ways in which students' family background may influence their rationale and motivations in making choices. There is evidence that students' socio-economic position could influence their choices directly, through the importance they place on intrinsic versus extrinsic benefits of study. For example, whether they want to choose a subject they personally enjoy and think they are good at, or whether they are more likely to consider labour market returns and outcomes upon graduation (Breen & Goldthorpe, 1997; Breen & Yaish, 2006).

The theory of relative risk aversion suggests students whose parents are better educated would be more inclined to choose subjects based on extrinsic motivations to avoid downward mobility, however, this study offers evidence to the contrary. Students whose parents are more educated are more likely to choose subjects based on intrinsic motivations. In line with Boudon's (1974) work, this could be due to the fact they are likely to have a 'safety-net' upon graduation, and compared to less advantaged students are more likely to succeed in whichever field they choose. In expanding the concept of primary and secondary effects of social background on education, Boudon argues that secondary effects, or educational choices, are driven by the different costs and benefits associated with these choices depending on background, which is demonstrated by the greater likelihood of securing a better paid job for more advantaged students, regardless of subject studied or university attended (Britton, Dearden, Shephard, & Vignoles, 2016).

There are a number of limitations to this study, which highlight a number of possible future research avenues. I have discussed theoretical concepts to help interpret results, including cultural and science capital, relative risk aversion and cultural reproduction theory, but I did not attempt to quantify these concepts. Whilst it may be argued, given the broad and relational nature of the concepts, that it would not be possible to do this, a number of studies have constructed quantitative measures of related concepts, including cultural capital (e.g. Van De Werfhorst & Hofstede, 2007; Zimdars, Sullivan, & Heath, 2009). Future researchers may wish to formalise these concepts and test how far they explain differences in subject studied by students' background. Furthermore, this study is interested in the impacts of students' different levels of enjoyment and perception of ability in subjects, but it is not within the scope of this chapter to test the reasons why these differences occur. Another fruitful area of research could be to test the extent that early streaming of students into subjects (Anders, Henderson, Moulton, & Sullivan, 2018; Iannelli & Duta, 2018), or allocation of teachers to different students, may influence students attitudes to subjects.

Chapter 7: The role of field of study in the transmission of socio-economic status from parents to children: Evidence from the UK and the US

7.1 Introduction

With the expansion of education systems across the world, increasing attention is being paid not only to how vertical differences in levels of education influence the transmission of social standing from parents to children, but also the implications of horizontal stratification within education (Charles & Bradley, 2002; Gerber & Cheung, 2008). Horizontal stratification occurs within each level of education (for example the subjects students choose to study) and have strong implications on later life outcomes (Altonji, Kahn, & Speer, 2016; Ishida, Spilerman, & Su, 1997; Sullivan er al., 2018; Webber, 2014). The extent that students are stratified into subject by gender is an ongoing and vibrant research area. Whilst research shows that students are also stratified into subjects by their social background, the patterns of stratification and its consequences are less well understood. This chapter aims to address this gap in the literature by comparing the relationship between social background and field of study for graduates in the UK and the US. Although these countries shared many similarities, there were a number of differences between them that may impact relationships. Firstly, in the UK students choose subjects very early compared to the US where students have more time to explore subjects equally before choosing one to focus on in depth. This may lead to greater stratification into subjects in the UK, where in the absence of a fuller exploration of subjects, students may choose subjects based on perceived ability, stereotypes and parental occupations. Secondly, although costs are converging now, university study was much more expensive in the US than in the UK, potentially causing young people in the US to be more concerned with returns after study when choosing a major, and this could be particularly true of students from less advantaged background (for whom the relative cost is higher). Thirdly, the expansion of university occurred earlier in the US, suggesting that students would have had more motivation to differentiate themselves along qualitative dimensions (field of study). Finally, research suggests that, although similarly low, income mobility was higher in the US than the UK (Bernardi & Ballarino, 2016).

Following research into differences in returns to subjects (Walker & Zhu, 2011), I define field of study by taking account of gender stereotyping, perceptions of relative difficulty of subjects, and expected occupational returns. The chapter then goes further by unpacking the implications of disparities in field of study in both countries on earnings in adulthood. Using longitudinal survey data from the BCS70 and the NLSY79 I am able to track individuals' occupational outcomes over time and compare differences and similarities by subject studied. This study aims to address the following main research questions:

1. How are students stratified by social background in fields of study?
2. To what extent do subject choices within university explain social background differences in later earnings?

The analysis presented in this chapter reveals broad similarities between the UK and the US in the ways students are stratified into subjects. In the UK, people studied similar subjects at university regardless of parents education, however women in the most disadvantaged groups (those who claimed Free School Meals at school), were most likely to study subjects which were both more lucrative and gender atypical. In the US, women whose parent's had lower education levels, but not at the very bottom of the income distribution, were more likely to choose these subjects. Robustness tests, which differentiated health and biological sciences from other Science, Technology, Maths, and Engineering (STEM) subjects, revealed that the association was not driven by less advantaged women sorting into more people oriented and female dominated STEM fields.

In both the UK and the US, taking account of field of study, rather than just level of education, did little to additionally explain differences in earnings differences by social background. This suggests that, at least in the time period of these studies, advantaged families were not (successfully) using field of study to maintain their income advantage.

This chapter proceeds as follows. First, I outline related literature and give an overview of the higher education contexts of the UK and the US, discussing differences that may be associated with sorting into subjects. Next, I outline the data used for analysis and construction of key variables. I go on to present descriptive analysis of differences in field of study by gender and family background, and differences in earnings by family background, field of study and gender. This sets the groundwork for my main regression analysis; first focusing on predicting field of study based on social background, next the extent that field of study explains earnings disparities. Finally, I conclude with a discussion of results and potential policy implications.

7.2 Literature review

7.2.1 The relationship between field of study and social background

With access to university expanding across countries, horizontal stratification within education is becoming a more active research area. This chapter focuses on one aspect of horizontal stratification within education, field of study. Compared to subject choices, other mechanisms for people to differentiate themselves within education depend more on prior achievement or financial means (i.e type of university attended). In comparison students are much less restricted in their choice of subject (although their choices are associated with achievement). The reasons and consequences for gender differences in subject choice has long been a vibrant area of research amongst social scientists across disciplines (e.g. Cech, Rubineau, Silbey, & Seron, 2011; England, Farkas, Kilbourne, & Dou, 1988). However compared to the study of gender disparities, differences in subject choices by social background are less clearly defined prior to this thesis. This is partly due to the difficulty of disentangling drivers of disparities. Less advantaged students have consistently lower performance than their more advantaged peers across subjects (Bukodi, Erikson, & Goldthorpe, 2014; Shavit & Blossfeld, 1993), and academic achievement independently predicts subject choices (The Royal Society, 2008).

The evidence that students choose different subjects depending on their families' characteristics in both the UK and the US has been extensively discussed in previous chapters and in the introduction of this thesis. In chapter five I studied the subject choices of young people born in 1989-1990, finding that students from less advantaged backgrounds were most likely to study a group of subjects that were more likely to offer high income returns upon graduation but weren't deemed as difficult as STEM subjects, including law, social science, business & management.

The most advantaged students were most likely to study arts and humanities subjects, which have lower returns on graduation. This research was timely, however the implications of these disparities on later-life earnings could not yet be tested. This chapter aims to address this gap in the literature using a sample of individuals born in 1970; who will have attended university more recently than the 1958 cohort but are also at an ideal age to assess implications for earnings in later life.

As discussed in previous chapters, the most prominent theories put forward to explain differences in educational choices by students' social background are Breen and Goldthorpe's (1997) theory of *Relative Risk Aversion*, and Lucas's (2001) theory of *Effectively Maintained Inequality.* The literature is mixed in support for these theories. Some studies suggest higher SES students are more likely to choose subjects with higher economic returns or prestige (e.g. Davies & Guppy, 1997), whilst results from chapter five suggest that more advantaged students are most likely to choose subjects which are less associated with high incomes or greater chances of employment, for example arts and humanities. This may be because they will likely succeed regardless of which subjects they study, and they have a 'safety net' in the form of parental support upon graduation. They will have parents with networks to help them succeed, and the financial resources to pursue unpaid internships or postgraduate study. They are likely more able to simply wait before entering the labour market, giving time to secure a more suitable job with better long-term prospects. They will also likely have the requisite social capital required to succeed in arts and humanities subjects (Van De Werfhorst & Hofstede, 2007), through time spent on extra-curricula activities outside of school or specific parenting practices that promote interest in (for example) literature or arts (Lareau, 2006).

7.2.2 Gender interactions

Chapters three and five outlined growing evidence that students' gender and social background interact in determining choices, with women particularly sensitive to SES effects on subject choice. For this reason, gender will also be a prominent aspect of analysis in this chapter.

The research pointed to a different hypothesised direction of associations than theories focusing on parenting styles and gender role attitudes. Women from low SES backgrounds may instead be *more* concerned about returns to study. All young people from less advantaged families are less likely to have the familial 'safety net' upon graduation discussed previously, however women may be particularly disadvantaged by this, because they are also disadvantaged in the labour market due to their gender. Thus, this combination of disadvantages may drive differences in choices.

The gendered aspect of subjects is also important. Charles and Bradley (2002) find that in more economically developed countries there is more gender segregation by subject (See also: Charles 2017; Sikora & Pokropek, 2012). They argue that in countries where the population are generally more advantaged, there is more opportunity for individuals to choose subjects they truly 'enjoy,' and affinity for specific subjects is likely to be influenced by gender stereotypes (Charles, 2011). The increased pressure for low SES women to choose a lucrative major may therefore overcome the influence of gender stereotypes within countries.

7.2.3 The relationship between field of study and later outcomes

It is not only important to understand the nature of stratification in education, but also it's effects on students' later life outcomes. The second aim of this chapter is to examine the extent to which horizontal stratification help to explain inequalities in graduate destinations (measured by earnings), and whether this differs in the UK and the US.

The reasons for associations may be direct or indirect. Some degrees may directly increase human capital more than others, for example through increased contact time or learning content, or they may act as a signal to employers that graduates have more skills. They could also simply promote skills that are more in demand in the labour market and thus have higher economic returns. On the other hand it could be that student stratification by subject area occurs independently of earnings differences. A high ability, high SES, young man may self select into typically high return subjects, but still have earned more regardless of subject studied. Thus, the observed differences in earnings by subject studied would actually be reflecting the different intake. This chapter uses the in depth information on students cognitive test scores to further test whether these differences are driven by disparities in prior attainment, or if they occur for similarly able young people. In the introduction of this thesis, I outlined the considerable body of research into the returns from specific subjects studied at university. Overall, the research suggested that this was some direct effect of field of study on earnings in adulthood, however this was reduced when controlling for family background and indicators of academic ability.

7.2.4 Field of study as a mechanism for intergenerational transmission of advantage

Less is known about the extent to which the effect of field of study on earnings explains differences in graduate earnings by level of advantage. Both the theory of *Relative Risk Aversion* and *Effectively Maintained Inequality* emphasise the importance of education in maintaining inequality, suggesting that if education were equal individuals should experience similar economic rewards upon graduation. In other words, education is the primary link between parent's attributes and their children's outcomes. The effect of social background that remains when accounting for education is the Direct Effect of Social Origin (DESO). For a discussion of DESO, and the research considering whether a direct effect really persists when controlling for education, see the introduction of this thesis. This chapter will test the extent to which DESO remains for university graduates when controlling not only for differences in performance in cognitive ability tests and prestige of university, but also field of study. Further, it will assess whether there are differences in the extent to which differences in field of study explain earnings disparities in the UK and the US.

7.3 Contextual features

Whilst there were many similarities between UK and US when study participants attended university, there were a number of key differences that could impact associations between social position, subject choices, and earnings. In both the UK and the US it was typical for students to begin their degree between 18-20 years old, and relatively less usual to study as a mature student. The subjects on offer were broadly similar, with some notable exceptions including medicine and law, which in the UK were offered as undergraduate degrees but in the US only as postgraduate degrees (however in the case of law, students in the UK would still have to study at postgraduate level to qualify as a practicing lawyer). Furthermore, degree length differed, with UK students typically studying for three years (with exception of Scotland, where students study for four years but are admitted a year earlier) and US students for four years.

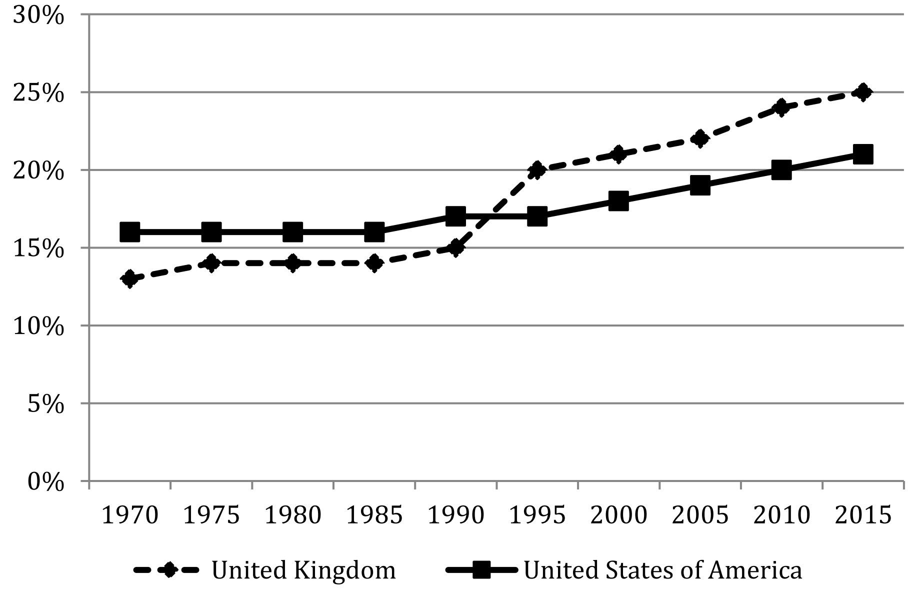
There are four distinctive features of the UK and US education systems that are particularly relevant to this study. The first is the timing of subject specialisation. In the UK, students were usually expected to specialise in subjects at a very early age. At 14 they make their first choices, however must continue to study core subjects maths, sciences, and English, and often at least one language. At 16 (two years before entry to university) they are expected to fully specialise in 3-4 subjects, and will be constrained in their choice of degree depending on the set of subjects they choose. Because many science subjects expect students to have studied at least two sciences, and the content of science qualifications often overlap or compliment each other, there were strong incentives to either entirely specialise in science and maths or to choose another route. Pre-university exams were subject specific. In contrast, in the US students study a much wider breadth of subjects before applying to university, and will typically apply to a university rather than a subject within a university (as is the case in the UK). University admittance in the US was based primarily on grade point averages and extra curricula activities, rather than grades in individual subjects. Whilst US students could gain admission partially based on grades in subject specific Advanced Placement exams, participation in these tests was voluntary and relatively rare for students in the 1970's and 1980's.

These differences in timing of specialisation are important because students expected to choose subjects earlier have less time to explore subjects equally and learn what they really enjoy. In lieu of a full exploration, they may choose subjects based on stereotypes of what they think is 'for' them, or what their family and peers have knowledge of. In terms of gender, there appears to be greater segregation of men and women into subjects in countries that differentiate students earlier (Charles and Bradley, 2002), and gender stereotypes are typically stronger in adolescence than early adulthood (Entwisle & Greenberger, 1972; Gaskell, 1984). Students from less advantaged backgrounds also face stereotypes about not being as ‘brainy’ as more advantaged children, and not as suited to ‘difficult’ STEM subjects (Archer et al., 2013; Campbell, 2015). Thus, it would be expected that less advantaged young people, particularly girls who face multiple negative stereotypes, would be less likely to study STEM subjects the UK than the US.

The second is the relative costs of study. Recently the cost of studying at university in the UK and the US has started to converge, but at the time the two study cohorts would have attended university differences were monumental. In the UK university was not only free, but living costs were either subsidised or fully covered, making the barriers to study beyond forgone wages relatively small. Students could receive means tested government grants to cover living expenses and were able to claim welfare benefits during their studies to cover housing. In contrast, in the US fees for study were substantial at public institutions, and even higher at private institutions. However, students from low-income families could apply for assistance and scholarships to help afford study, and loans were available to cover costs (often at high rates of interests). Nonetheless, the cost of study may have influenced students’ choices of field of study. In the US, Kane (1994) showed that increased university fees decreased participation of less advantaged students, and access to larger grants increased participation (Dynarski, 2003). In contrast, Sá (2017) finds that whilst participation of students generally fell after fee increases in the UK, this was not more pronounced for less advantaged students, and all students were more likely to study more lucrative fields after fee reforms. Thus, it may be expected that the presence of fees in the US to impact field of study, particularly for students from disadvantaged backgrounds, for whom the relative cost of study (and potential debt) is higher.

Third, whilst both countries experienced a large expansion of university systems in the latter part of the twentieth century, this expansion occurred at different rates. Figure 7.1 shows that at the time study participants would have attended university, young people in the US were more likely to have post secondary qualifications than young people in the UK. The expansion of universities appears to have occurred earlier in the US, and proportions of students with post-secondary qualifications in the UK overtake the US around 1990. This indicates that students in the US may have started to differentiate themselves by field of study earlier, and that any social differences would be stronger for the US cohort.

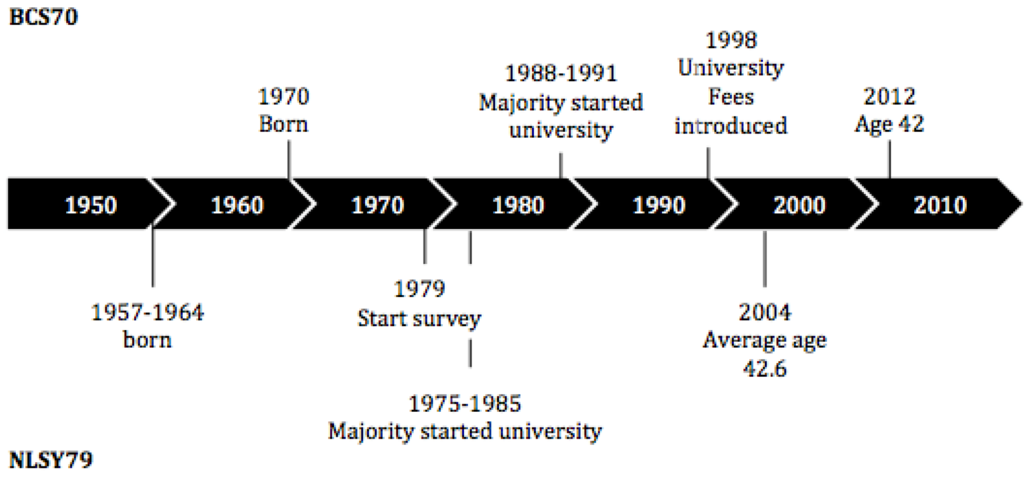
Figure 7.1: Share of the population aged 20-24 with post-secondary education between 1970-2015[1]



Finally, another area of possible difference between the UK and the US is the extent of earnings differences by family background both for individuals who did, and did not attend university. In countries where there is a larger association between background and outcomes in adulthood, particularly amongst graduates, we may expect students to be more likely to use subject choice as a vehicle for social mobility. As noted in the introduction, some research suggests the relationship is larger in the UK than the US even when controlling for education (Bernardi & Ballarino, 2016).

[1] Data from: http://www.oeaw.ac.at/fileadmin/subsites/Institute/VID/dataexplorer/index.html

Figure 7.2: Timeline of lifecourse and policy events for the BCS70 and NLSY79 cohorts



7.4 Data and variables

7.4.1 BCS70 data

The British Cohort Study (BCS70) is a longitudinal birth cohort study that started in 1970 following just over 17,000 people born across Great Britain. Data were collected when individuals were 5, 10, 16, 26 and every 4 years thereafter. In the most recent wave of data available for analysis, collected in 2012, individuals were 42 years old. Data on a wide variety of characteristics and outcomes were collected including social circumstances, cognitive ability, educational achievement and choices, and occupational outcomes. Of a total sample of 9,448 individuals, 1,968 obtained a degree by 42 and recorded gender and subject choice.

The majority of individuals who attended university in this cohort would have started their courses in the late 1980's – early 1990's. This was after the great expansion of the higher education system that followed the Second World War, just before (or during) the conversion of polytechnics[1] (which traditionally offered more vocationally oriented degrees) to full university status, and before the introduction of tuition fees (initially in 1998) and erosion of financial assistance for students. There were also large differences in the number of students attending

[1] Graduates from polytechnics are included in the study because they were likely to have converted to universities by the time cohort members graduated.

university in this time, compared to the current day, with around one in five 18 year-olds attending university in 1990. In the BCS70 sample 23.5% of the total 2012 cohort had attained a degree by 42.

Multiple imputation using chained equations was employed to account for missing data on other variables, with 20 datasets created (White, Royston, & Wood, 2011). For full specification of variables used in imputations, see appendix A. Because this accounts for missing data, but not attrition over time, longitudinal weights were also constructed using logistic regression modelling predicting probability of being in the most recent wave based on baseline characteristics. Characteristics chosen were informed by Mostafa & Wiggins (2014), and included sex, birthweight, parity, mother's age, whether mother lived in the southeast of England in the first survey, social class at birth, and mother's and father's age at completion of education.

7.4.2 NLSY79 data

The National Longitudinal Survey of Youth 1979 (NLSY79) is also a longitudinal panel study, in which participants were interviewed annually up to 1994 and biannually thereafter. The survey began with a nationally representative sample of 12,686 young people, and covers a broad range of topics and outcomes, with detailed information on education, cognitive ability and occupational outcomes. Unlike BCS70 the study is not a birth cohort. Individuals were between 14-22 at the time of the first survey, and were born between 1957-1964. Thus, although the cohorts are close in age, there is some time lag, and additional age controls for the NLSY79 cohort were included in analysis. If they attended university between 18 and 21 years old, they would have attended between the years 1975-1985. I restricted the sample to those who were dependent on their parents at the start of the study, as only for these participants did parents complete questions about income, and to those who were 21 or under. The final sample included 7,856 individuals, of which 1,571 were graduates who had completed a four-year degree by 2004 (26% of the sample), with a mean age[1] of 16.8 at the start of the survey, and 42.5 by 2004.

[1] Reported ages were adjusted using sampling weights provided in NLSY79

7.4.3 Key variables

Field of study

Subjects were grouped into three categories including: Science, Technology, Engineering and Maths (STEM) subjects; Law[1], Economics and Management (LEM) subjects, and Other Social Sciences, Arts and Humanities subjects (OSSAH). Whilst historically many studies have classified field of study as a binary one between STEM and non-STEM subjects, a number of studies suggest that including a further group of subjects with positive occupational outcomes (LEM subjects) gives a better overall picture of differences in outcomes (e.g. Walker and Zhu, 2013). Students who studied a joint-honours degree (unless they had a primary subject) were classed as OSSAH. The majority of people had studied an OSSAH subject, followed by STEM subjects, then LEM subjects. Overall, more people in the UK sample had studied an OSSAH subject compared with the US sample, and in the US more people had studied a LEM subject.

Table 7.1: Proportion of graduates holding a degree in each subject group in BCS70 and NLSY79

Field of study	UK (BCS70)		US (NLSY79)	
	N	%	N	%
STEM	633	32.1	503	31.6
LEM	373	18.9	395	25.5
OSSAH	962	49.0	673	42.9
Total	1,968	100	1,571	100

Family background

Parents' education was the primary measure of family background used in this study. In BCS70, highest qualification levels were recorded, and in NLSY79 years of study within educational levels were recorded. I have attempted to match relative levels of education with qualifications (shown in table 7.2). In the final measure, parents who studied in post-compulsory education are classified as 'highly educated.' In both cases the highest qualification of both parents, or the

[1] Whilst 'law' itself was not included in the US sample, pre-law and related subjects were (including 19 individuals). These were considered equivalent to law in the UK, as in both cases further study would be required to becoming a practicing lawyer, and would not automatically lead to a career in law.

qualification of the only parent, was used for analysis[1]. Overall parents of degree holders in the US sample are more educated than the UK sample.

A number of other family background characteristics were included, such as eligibility for Free School Meals (in the UK), family income (in the US), whether participants had attended a private (fee paying) school, and whether they grew up without a biological or step father in the household (in the US this is measured at the first survey, in the UK at any point between 5, 10, and 16 years old). Ethnicity is also included in models, in the UK people are classified as white or Black and Minority Ethnicity (BME), due to low ethnic variation, and in the US people are classified as black, Hispanic, or non-black, non-Hispanic.

Table 7.2: Family background measures in the BCS70 and NLSY79

	Classification		Whole sample (%)		Graduates (%)	
	UK	US	UK	US	UK	US
High parental education	A levels or above	Some college or above	19.2	33.1	52.7	63.2
Low parental education	O-levels or below	High school or below	80.8	66.9	47.3	36.8
Low income/ poverty	Free school meals	Bottom 10% income	13.3	11.1	4.1	4.3
Privately educated			1.5	5.8	8.4	11.6
Father (bio or step) not in household	At 5, 10 or 16 (any)	At first interview	18.4	21.5	11.2	15.2

Earnings

Earnings for the BCS70 sample were computed from two variables: Gross earnings from work, and period earnings cover (for example weekly, monthly, yearly). Individuals whose period of earnings was not disclosed had to be coded as 'missing.' Yearly earnings for the NLSY79 sample used the truncated gross income from salary and wages. In both countries measures were taken when participants were 42 (on average for the NLSY79 sample). Yearly earnings were then converted to 2017 prices using CPI figures.

[1] In the majority of cases, the father's education was higher than the mother's education. Robustness checks included parent's education separately in specifications.

Gender

Models are run separately for men and women, because students' gender strongly predicts the subjects they will choose to study throughout education. Girls and women are most likely to choose subjects that are people-focused and require greater language and communication skills (typically arts, humanities and health spheres), whilst male students are more likely to choose subjects with strong mathematical or technical components (Cech et al., 2011; England et al., 2007). Whilst this gendered pattern of subject choice is a persistent phenomenon across countries and time-points, there are differences in the extent of disparities across countries (Charles & Bradley, 2002).

Cognitive tests

For the UK measure a composite score was created using Principle Components Analysis (PCA), combining scores in age 5 and 10 tests including; the copying designs test, the human figure drawing test, the English Picture Vocabulary Test (EPVT), the complete a profile test, and the Schonell reading test, Edinburgh Reading Test and the Friendly Maths Test. Cronbach alpha for these items was 0.75. For more information on cognitive ability measures provided in BCS70 see Parsons (2014). Cognitive ability in NLSY79 was measured using scores from the Armed Forces Qualification Test (AFQT) taken around the beginning of the survey, calculated using 2006 standards. Both measures were standardized to mean zero, standard deviation one. Because the US measure of cognitive ability was taken in adolescence/ early adulthood, it is likely this measure will be more strongly associated with differences in upbringing (therefore more reflective of social background). It is also likely the case that both measures are highly reflective of circumstance, and should be interpreted as measures of achievement in a low stakes test rather than 'innate' ability.

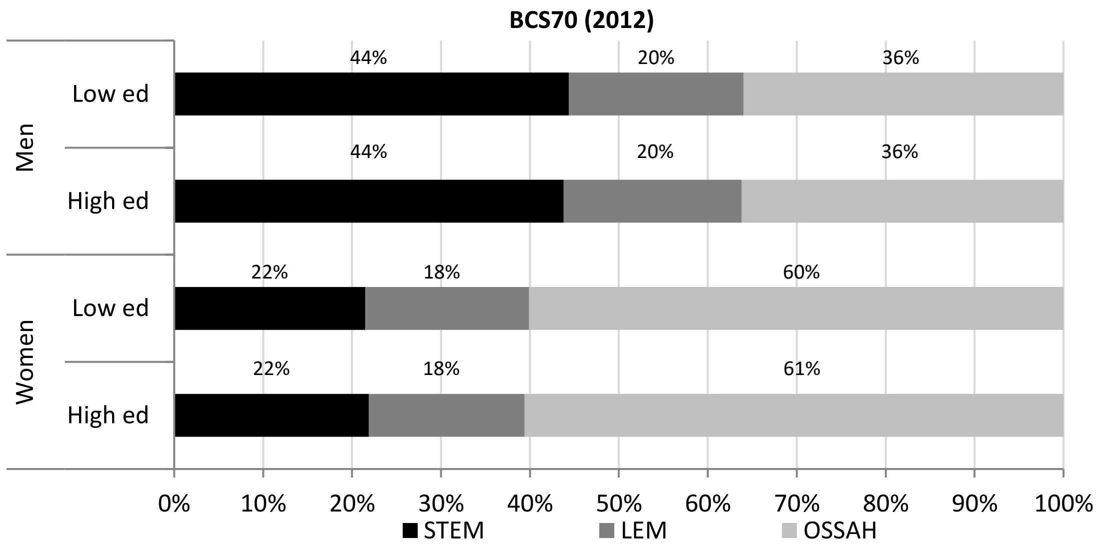
University prestige

University type is strongly associated with outcomes after university, but is also associated with the subjects people choose to study (Bostwick, 2016); students may choose an easier (or less subscribed) subject to gain acceptance to a higher-ranking university. Because the returns to degrees by university are extremely varied, the decision was made to include a continuous rather than categorical variable. Average UCAS points[1] (pre university exam scores) of admitted students were linked with the university. This acts as a measure of the competition for places at the university (for example, the competition for places at Oxbridge is much higher than other prestigious universities, and whilst universities may have similar entrance requirements their admitted students may have very different final UCAS scores), a proxy for individuals' academic achievement, and a further signal to employers of higher ability.

7.5 Descriptive results

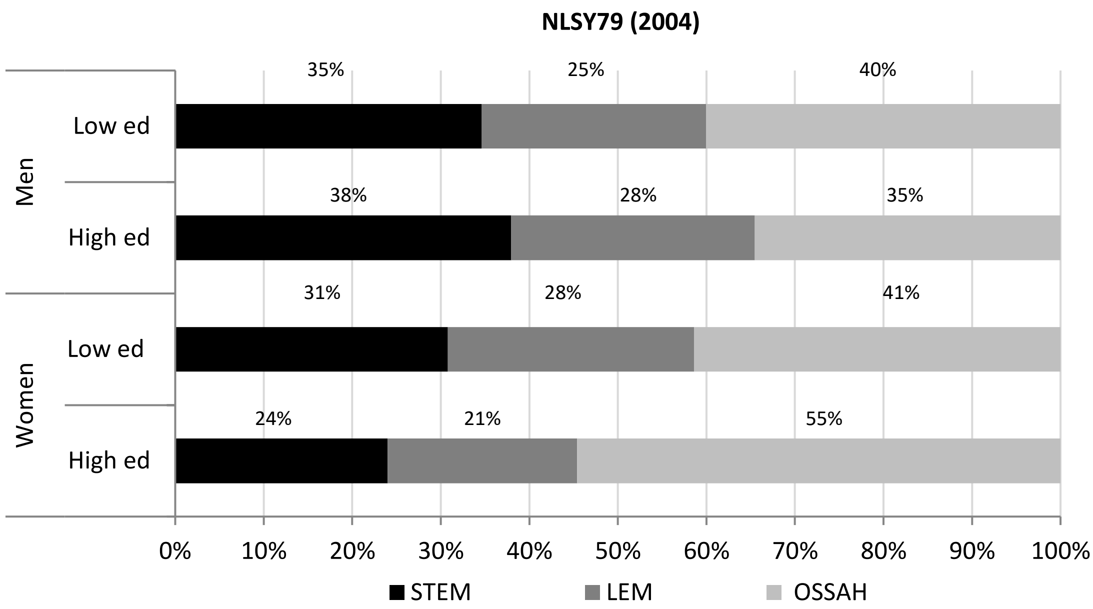
Figures 7.3 and 7.4 chart the field of study of UK and US graduates respectively by their parents' education level and own gender. In the UK women are more likely to have studied OSSAH than men, and much less likely to have studied STEM. Whilst the overall gender differences are similar in the US they are less pronounced.

Figure 7.3: Subject studied by parents' education and gender in the BCS70 sample



[1] Students could gain UCAS points from A-level qualifications (or equivalent), and from some extra curricula activities (i.e. high level music qualifications).

Figure 7.4: Subject studied by parents' education and gender in the NLSY79 sample



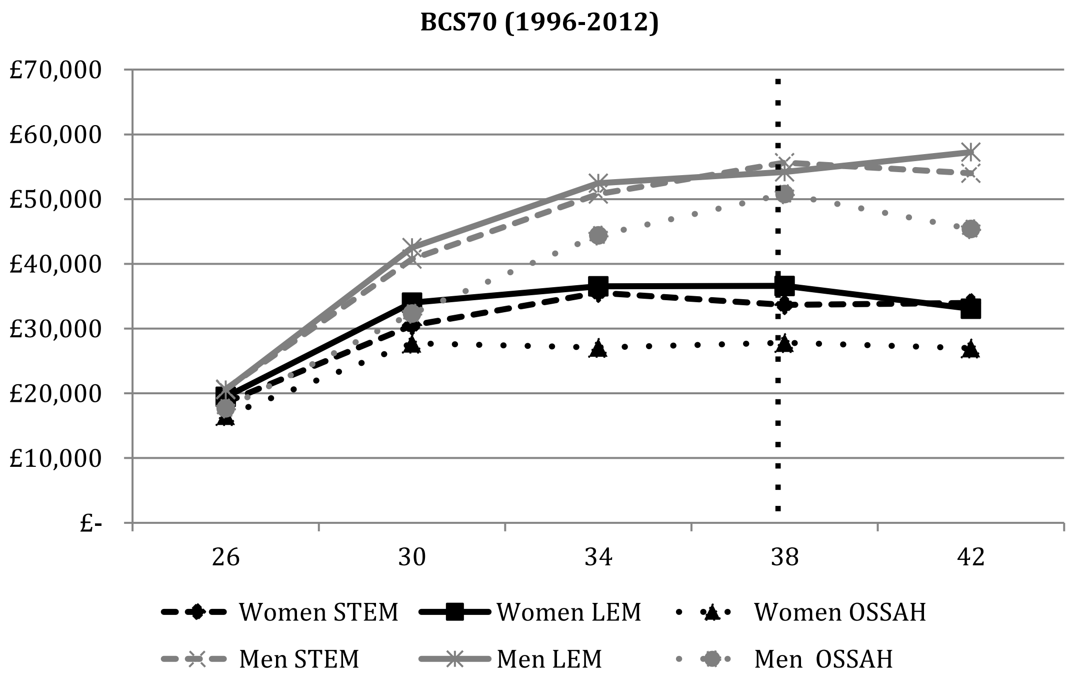
In the UK differences in subject studied by parents' education level are very small for both men and women. In contrast, in the US whilst there is very little association between parent's education and subject studied for men, there are substantial differences in choices for women. Women with more educated parents are least likely to have studied either a STEM or LEM degrees. The largest differences are seen in participation in OSSAH subjects; women with more educated parents are 14 percentage points more likely to have an OSSAH degree than women with less educated parents. This is in line with Ma's (2009) finding using NELS data that higher SES women were less likely to choose traditionally male dominated subjects. The UK findings contrast with findings from the UK and the Netherlands for more recent graduates, suggesting that higher SES women are more likely to choose STEM subjects (Codiroli Mcmaster, 2017; Van De Werfhorst, 2017).

7.5.1 What's the relationship between field of study and earnings trajectory?

For UK graduates earnings trajectories are characterised by a sharp rise around the beginning of their careers and a levelling out thereafter, whereas in the US graduate earnings appear to rise more steadily across their careers. Median earnings are substantially higher for men than women throughout both the BCS70 and NLSY79 periods, and women's yearly earnings plateau at an earlier age than men's. In both the UK and the US, graduates who studied OSSAH subjects earn consistently less than those who studied either STEM or LEM subjects, however whilst in the UK STEM and LEM graduates had remarkably similar median earnings, in the US STEM graduates earn more, particularly after age 42. STEM graduates (and LEM graduates in the UK) appear to have recovered quicker from the recession, in line with recent findings from NLSY79 (Altonji et al., 2016). In the US, differences in earnings by subject studied are much smaller for women compared to men.

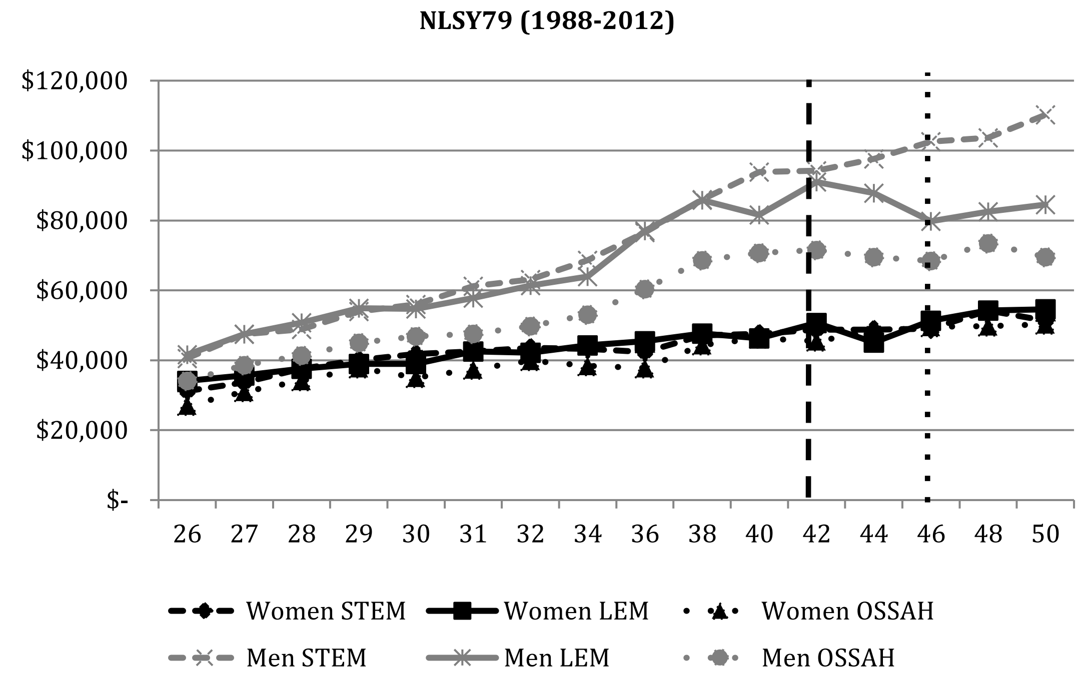
It is possible differences are explained (or suppressed) by differences in the students who study each group of subjects. For example, the social profile of students who typically study a group of subjects may explain earnings differences, rather than the subjects themselves having higher returns. This could go in the other direction; the earnings differences between people from different social backgrounds may be either exaggerated or suppressed by the subjects they sort into at university.

Figure 7.5: Median yearly earnings by subject of degree and gender in the BCS70 sample[1]



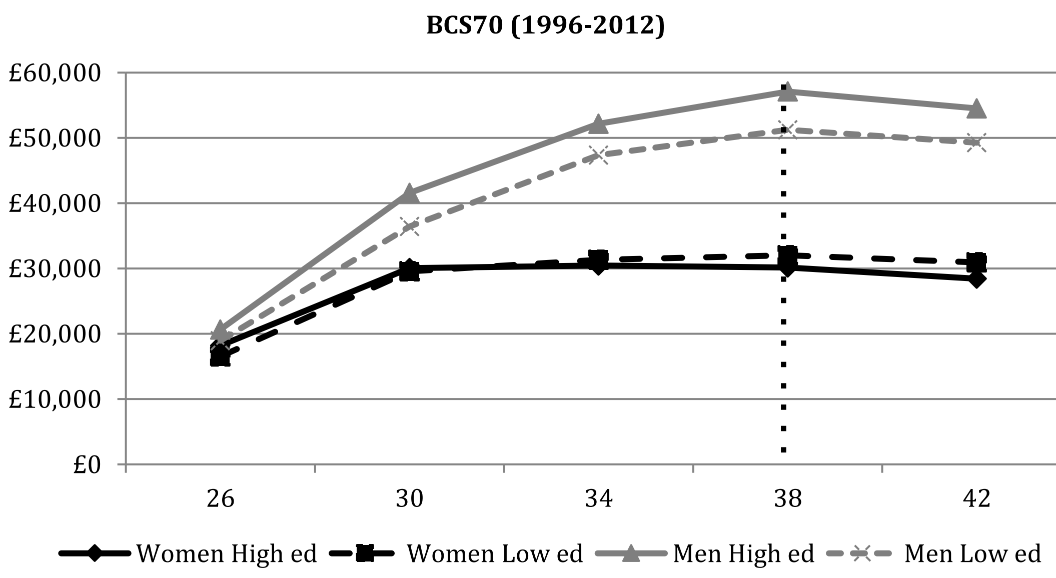
[1] The vertical dotted line shows where the recession took place (2008).

Figure 7.6: Median yearly earnings by subject of degree and gender in the NLSY79 sample[1]



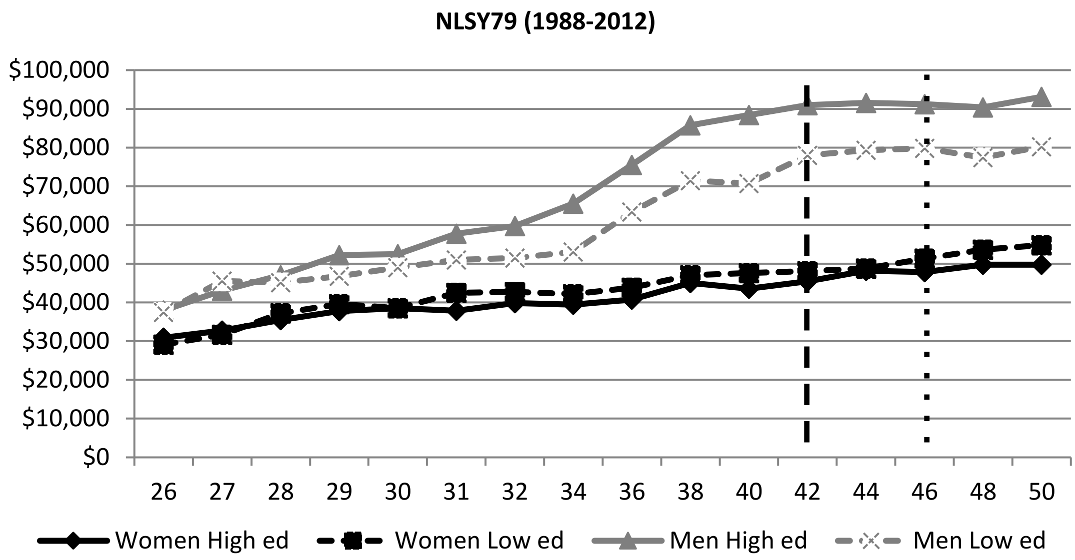
7.5.2 *What are the differences in earnings trajectories by family background?*

Figure 7.7: Median yearly earnings by parents' education and gender in the BCS70 sample



[1] The vertical dotted shows where the recession took place (2008), and the vertical dashed line indicates where the NLSY79 sample were the same age (42) as the people in the most recent data collected in BCS70.

Figure 7.8: Median yearly earnings by parents' education and gender in the NLSY79 sample



Whilst differences in earnings by family background were smaller than gender differences, they are still substantial. The differences are starker for men than for women, with men whose parents are more educated earning more than those whose parents had lower levels of education.

7.6 Multivariate results

7.6.1 How are students stratified by social background in fields of study?

To test whether differences in field of study are statistically significant when controlling for other factors related to the subjects people choose, three multinomial logistic regressions were run. The regressions were built up in stages. The first model included just student's parents' education, their gender and their ethnicity, to estimate the raw differences in choices by background characteristics. The second model additionally included whether they claimed Free School Meals in the UK, whether they were in the bottom 10% income decile in the US, or whether they attended a private school. Models one and two were run separately to test whether inclusion of these additional family background controls confounded results for parental education, given the high overlap across measures. The third model additionally controls for individuals cognitive ability test scores, which is associated with both field of study (particularly studying STEM subjects) and family circumstances. The UK analysis includes a fourth model that controls university prestige. Thus the underlying regression equation is:

$$y_i = \alpha_i + \beta_1 FB_i + \beta_2 R_i + \beta_3 A_i + e_i$$

Where *y* is the probability of having a STEM or LEM degree over a OSSAH degree, or ln[prob(STEM or LEM)/prob(OSSAH)]. FB is the students' family background, R is the students Race or ethnicity, and A is ability. The UK regressions additionally include university status, and the US regressions include age controls to account for the fact that participants were interviewed at different ages.

Table 7.3 shows results for women in the UK and the US. The regression results in all models broadly reflect the descriptive statistics. In the UK there was no association, either statistically or substantively, between parents' education and field of study. In contrast in the US, women whose parents had not attended college were around twice as likely to choose STEM or LEM instead of OSSAH subjects in model 3, compared with women whose parents had higher education levels. These women appear to be less inclined to study less lucrative, 'feminine,' OSSAH subjects.

In the UK, women who had claimed Free School Meals were more likely to study both STEM and LEM subjects (shown in model 2), suggesting a similar association with the US for the most disadvantaged young women. As FSM eligibility is by definition linked with family income, a binary variable ' lowest decile income' was included in the US regressions, which as shown in the descriptive statistics contained a similar proportion of students who attended university[1]. This was, however, not significantly associated with field of study.

[1] Robustness tests were run which included a continuous income predictor in the US models that was not statistically significant. Ideally, a similar 'bottom decile' variable would have been constructed, however in BCS70 family income was grouped with fewer levels and was not significant when included in specifications. 'Living in council housing' was also considered as a variable included in both surveys as a measure of economic disadvantage, however the US sample contained an extremely small number of graduates who had lived in social housing (1.32%) compared to the UK sample (11.55%). In the UK, living in council housing was associated with choice of STEM subjects over OSSAH, but not LEM subjects, for both men and women.

Table 7.3: Multinomial logistic regression predicting field of study for women

	UK								US					
Base subject: OSSAH	STEM				LEM				STEM			LEM		
	1	2	3	4	1	2	3	4	1	2	3	1	2	3
Social background														
Low parental education	0.952	0.921	0.973	1.044	0.970	0.986	0.970	0.946	1.680**	1.716**	1.848***	1.695**	1.655**	1.791**
	(0.158)	(0.155)	(0.168)	(0.183)	(0.176)	(0.185)	(0.185)	(0.182)	(0.375)	(0.387)	(0.430)	(0.383)	(0.380)	(0.418)
Free School Meals/ Lowest decile income		1.843	2.023	2.006		2.500**	2.428*	2.449*		0.813	0.850		0.575	0.596
		(0.859)	(0.954)	(0.972)		(1.159)	(1.129)	(1.133)		(0.413)	(0.439)		(0.338)	(0.347)
Independent School		0.549	0.533*	0.447**		1.264	1.274	1.363		1.685	1.680		0.987	0.979
		(0.202)	(0.197)	(0.163)		(0.404)	(0.406)	(0.448)		(0.604)	(0.608)		(0.418)	(0.416)
Father (bio or step) not present		0.738	0.748	0.774		0.641	0.641	0.631		1.124	1.211		1.846*	2.003**
		(0.212)	(0.216)	(0.224)		(0.205)	(0.205)	(0.202)		(0.355)	(0.387)		(0.578)	(0.635)
Ethnicity														
BME	1.374	1.254	1.424	1.458	1.932	1.742	1.679	1.665						
	(0.718)	(0.661)	(0.763)	(0.788)	(1.101)	(1.001)	(0.984)	(0.967)						
Reference: Non-black, non-Hispanic														
Hispanic									1.827*	1.802*	2.233**	1.726	1.605	2.010*
									(0.597)	(0.596)	(0.766)	(0.604)	(0.565)	(0.723)
Black									1.636**	1.656**	2.318***	1.663**	1.559*	2.202**
									(0.381)	(0.411)	(0.695)	(0.406)	(0.412)	(0.679)
Cognitive test score (sd)			1.185	1.149			0.953	0.963			1.394**			1.414**
			(0.126)	(0.124)			(0.107)	(0.109)			(0.231)			(0.236)
Prestige of university														
Mean university acceptance scores				1.004***				0.999						
				(0.001)				(0.001)						
Age									0.140	0.233	0.320	2.634	2.050	2.929
									(0.282)	(0.473)	(0.657)	(5.622)	(4.445)	(6.363)
Age^2									1.022	1.016	1.012	0.988	0.991	0.987
									(0.024)	(0.024)	(0.024)	(0.025)	(0.025)	(0.025)
Observations	1054								839					

* p < 0.10, ** p < 0.05, *** p < 0.01

A measure that could go some way to explaining differences in results is family type, and whether the individuals grew up without a father present. This is likely to be associated with both parental education in both countries, and Free School meal receipt in the UK (because this is based on benefit receipt, children in lone parent families are more likely to be eligible than those in two parent households with similar income). Women without a father figure in the home would have had a female role model taking primarily responsibility for finances, therefore may be even more concerned with financial returns after study. People without a father in the household are also less likely to have a parent who worked in STEM, or a strong interest in STEM. Thus, it may be expected that people without a father were less likely to study STEM. For the UK sample having no father present in the household had no statistical association with field of study. In the US, not having a father present was positively associated with studying LEM over OSSAH, but not studying STEM. This is in line with predictions; they were choosing more lucrative subjects, but not more likely to study STEM (as a consequence of lower likelihood of having a parent with an interest in, or working in, STEM). Importantly, the addition of 'lone parent' to the model does not change coefficients on parents' education. This suggests US women in two parent families are also more likely to study STEM/ LEM if their parents are less educated.

There are a number of other interesting exploratory findings relating to the control variables. In the US women with higher ability were more likely to study STEM or LEM subjects over OSSAH, however in the UK there was no association between field of study and ability. This could be due to timing of assessment, or differences in the measures of testing. In the UK, mean university acceptance scores, which are likely reflective of the individuals own academic achievement prior to university entry, were associated with STEM study. It is also the case that high prestige universities are more likely to offer STEM courses. BME women in the UK were not more likely to study STEM or LEM, and in the US Black women were more likely to choose STEM or LEM over OSSAH. It is important however to remind the reader that being a birth cohort, the BCS70 does not include the choices of individuals not born in the UK. It should also be noted that 'BME' in the UK contexts includes many individuals who in the US would be classified as non-black non-Hispanic, for example Asian individuals.

Table 7.4: Multinomial logistic regression predicting field of study for men

	UK								US					
Base subject: OSSAH	STEM				LEM				STEM			LEM		
	1	2	3	4	1	2	3	4	1	2	3	1	2	3
Social background														
Low parental education	0.991	0.925	1.001	1.037	0.971	0.906	0.891	0.877	0.796	0.791	0.887	0.762	0.780	0.772
	(0.157)	(0.154)	(0.171)	(0.179)	(0.197)	(0.190)	(0.191)	(0.190)	(0.184)	(0.187)	(0.215)	(0.197)	(0.204)	(0.210)
Free School Meals/ Lowest decile income		1.513	1.701	1.665		1.260	1.232	1.242		1.477	1.778		0.372	0.376
		(0.739)	(0.834)	(0.815)		(0.737)	(0.722)	(0.728)		(0.948)	(1.071)		(0.314)	(0.314)
Independent School		0.536**	0.511**	0.474**		0.673	0.678	0.699		0.804	0.802		1.317	1.316
		(0.164)	(0.157)	(0.148)		(0.243)	(0.245)	(0.256)		(0.302)	(0.301)		(0.480)	(0.480)
Father (bio or step) not present		0.795	0.780	0.803		1.149	1.155	1.142		0.271***	0.249***		0.743	0.745
		(0.244)	(0.243)	(0.252)		(0.390)	(0.392)	(0.387)		(0.098)	(0.091)		(0.246)	(0.248)
Ethnicity														
BME	2.481*	2.411*	3.197**	3.151**	1.091	1.086	1.022	1.027						
	(1.189)	(1.173)	(1.603)	(1.583)	(0.757)	(0.764)	(0.743)	(0.747)						
Reference: Non-black, non-Hispanic														
Hispanic									0.696	0.748	0.911	0.841	0.861	0.841
									(0.245)	(0.271)	(0.355)	(0.346)	(0.353)	(0.353)
Black									0.855	1.063	1.709	0.741	0.891	0.836
									(0.243)	(0.335)	(0.638)	(0.236)	(0.295)	(0.321)
Cognitive test score (sd)			1.298***	1.275**			0.951	0.958			1.613**			0.948
			(0.129)	(0.129)			(0.115)	(0.117)			(0.310)			(0.169)
Prestige of university														
Mean university acceptance scores				1.001				0.999						
				(0.001)				(0.001)						
Age									4.633	4.577	2.361	1.039	1.107	1.168
									(9.903)	(9.888)	(5.182)	(2.294)	(2.480)	(2.626)
Age^2									0.983	0.983	0.990	1.001	1.001	1.000
									(0.024)	(0.025)	(0.025)	(0.026)	(0.026)	(0.026)
Observations	914								732					

* p < 0.10, ** p < 0.05, *** p < 0.01

Table 7.4 illustrates the equivalent results for men. As with the descriptive statistics, parental education was not associated with field of study for men. There are a number of reasons why this association may only be observed for women. They face a double disadvantage in the labour market, and may make decisions accordingly by placing higher value on returns after education than men from similar social backgrounds. Furthermore, for men being from a low educated family and being more motivated by economic returns wouldn't be at odds with choosing gender-typical (STEM) subjects, and may therefore have less influence over choices. For women, pursuing more lucrative subjects would generally mean choosing gender atypical subjects, and perhaps making very different choices than they would if they were less concerned about future outcomes. In the UK not having a father was not associated with field of study, however in the US men who grew up without a father were less likely to study STEM. This finding is in line with the previous interpretation for women; men do not get the same gender 'role model' effect of seeing a women support her family, but more likely suffer a lack of a 'STEM' role model.

In terms of control variables, men with higher ability in both the UK and the US were more likely to study STEM. In the UK, BME men were more likely to study STEM, but in contrast to findings for women, there was no association between race and field of study for US men.

7.6.2 To what extent does horizontal stratification in university explain social background differences in later earnings?

To address the second aim of this chapter; to test the extent that earnings disparities are explained by differences in subject choices, a series of regressions predicting log earnings at age 42 were run. Total earnings from salary and wages per year was chosen as the outcome of interest, rather than hourly earnings, because working part time was thought of as a mechanism through which inequalities in earnings can manifest[1]. For example, people who grew up in more advantaged households may have better access to jobs with predictable working hours, and may be less likely to have spent time out of work over the year.

[1] Heckmen two-stage regressions were also run, selecting firstly on probability of being economically active (employed or unemployed), and secondly on probability of being unemployed, to test whether results were similar.

These regressions were built up in three stages, the first included just demographic characteristics, the second included an indicator for whether they had a degree at all or not, and the third included additional levels of field of study to the education indicator. The extent that field of study mediates differences in outcomes by social background, over and above just educational level, is calculated as the coefficient on parents' education in model 3 minus the coefficient on model 2. These models were run for men and women separately. The underlying regression equations are:

$$log(earnings)_i = \alpha_i + \beta_1 FB_i + \beta_2 R_i + \beta_3 A_i + \beta_4 Degree_i + e_i \text{ or}$$

$$log(earnings)_i = \alpha_i + \beta_1 FB_i + \beta_2 R_i + \beta_3 A_i + \beta_4 Degree\_subject_i + e_i$$

Results outlined in tables 7.5 and 7.6 show that parents education is strongly associated with earnings by 42 in both the UK and the US, however the raw effect appears larger in the UK. This is likely at least partially driven by the fact that parents in the US were more likely to have some post compulsory education than parents in the UK. In the UK, controlling for whether participants obtained a bachelors degree reduced the differences in earnings by parent's education. However, degree attainment appeared to have a larger effect on disparities in the US; differences were no longer significant for women, however remains substantial for men.

Table 7.5: Differences log earnings by family background – women (quantile regression models)

	UK			US		
	Model 1	Model 2	Model3	Model 1	Model 2	Model3
Parents have high education	0.378***	0.159***	0.154***	0.283***	0.0599	0.0556
	(0.0357)	(0.0391)	(0.0404)	(0.0794)	(0.0744)	(0.0760)
Degree		0.693***			0.641***	
		(0.0369)			(0.0665)	
Subject studied						
No degree						
STEM			0.796***			0.552***
			(0.0705)			(0.134)
LEM			0.848***			0.736***
			(0.0953)			(0.123)
OSSAH			0.613***			0.644***
			(0.0437)			(0.0737)
Ethnicity						
BME	0.241**	0.229**	0.229***			
	(0.0967)	(0.0936)	(0.0876)			
Reference: Non-black, non-Hispanic						
Hispanic				0.0469	0.104	0.0914
				(0.0977)	(0.0831)	(0.0835)
Black				-0.0641	-0.0553	-0.0607
				(0.0716)	(0.0642)	(0.0631)
Age				-0.0864	0.417	0.324
				(0.744)	(0.673)	(0.685)
Age^2				0.000955	-0.00487	-0.00376
				(0.00867)	(0.00782)	(0.00796)
Constant	9.580***	9.509***	9.510***	12.04	1.066	3.055
	(0.0231)	(0.0232)	(0.0239)	(15.96)	(14.46)	(14.73)
N		5,120			4,205	

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

In the UK, women appear to gain a higher earnings premium after obtaining a degree than men, however in the US men and women are similar in their gains from a bachelors degree. There is also considerable variation in earnings by field of study; graduates in both countries earn a much higher premium over non-graduates after studying LEM and STEM subjects than OSSAH subjects. Despite the considerable differences in returns by subject studied, including this additional information compared to just 'obtained a degree' did little to explain differences in earnings by parents' education. Whilst these differences are not statistically significant, if anything a small increase in disparities in earnings is observed when accounting for subject choices. Results run contrary to the *effectively maintained inequalities* hypothesis, that is, field of study was not a mechanism through which more advantaged individuals enhanced their advantage in later earnings. Given that less advantaged US women appeared to study subjects associated with higher returns, we may have expected differences in earnings by social background to increase to a much larger extent when controlling for subject studied, however differences remain remarkably similar. This suggests that, although higher SES women studied less lucrative majors, they were able to make up for any disadvantages this may bring.

Table 7.6: Differences in log earnings by family background – men (quantile regression models)

	UK			US		
	Model 1	Model 2	Model3	Model 1	Model 2	Model3
Parents have high education	0.313***	0.161***	0.170***	0.343***	0.159***	0.167***
	(0.0325)	(0.0292)	(0.0300)	(0.0468)	(0.0433)	(0.0413)
Degree		0.480***			0.644***	
		(0.0306)			(0.0483)	
Subject studied						
No degree						
STEM			0.540***			0.697***
			(0.0374)			(0.0526)
LEM			0.602***			0.739***
			(0.0615)			(0.0712)
OSSAH			0.355***			0.495***
			(0.0434)			(0.0691)
Ethnicity						
BME	0.0735	0.0267	0.0147			
	(0.106)	(0.103)	(0.100)			
Reference: Non-black, non-Hispanic						
Hispanic				-0.236***	-0.180***	-0.183***
				(0.0507)	(0.0535)	(0.0515)
Black				-0.605***	-0.551***	-0.545***
				(0.0698)	(0.0664)	(0.0641)
Age				-0.658	-0.563	-0.547
				(0.446)	(0.376)	(0.340)
Age^2				0.00774	0.00683	0.00660*
				(0.00520)	(0.00437)	(0.00395)
Constant	10.30***	10.24***	10.24***	24.74***	22.27***	22.00***
	(0.0152)	(0.0146)	(0.0148)	(9.562)	(8.081)	(7.293)
N		4,354			4,622	

* p < 0.10, ** p < 0.05, *** p < 0.01

7.6.3 Heckman selection models

Following Britton et al (2016), earnings from salary and wages are used in analysis, as opposed to overall income or family income. If an individual is not in employment they have zero earnings. This reflects the real earnings differences across individuals. Where an individual is not in work, for example if they are unemployed, this could be seen as a mechanism through which income disparities by either family background or education emerge, for example if subject studied was associated with access to employment (Altonji et al., 2016). Therefore, if the regressions were restricted only to measure earnings of people in employment, the coefficients would not reflect all ways that a person's background and education may influence earnings.

It may be argued, however, that including income as zero for *all* people out of work, including those who stay at home to look after their children, may bias results for women. The main problem here would be the pooling of negative states, for example, women who choose to take time out of work to look after young children, and those who cannot find work, are coded as having the same income. Many women who choose not to work will be sharing income with a partner, who may work longer hours as a consequence. Thus, the woman at home has zero earnings in the data, but is contributing in other ways to household earnings, which she then benefits from. Particularly, if family background were associated with higher likelihood of staying at home with children, the association between background and earnings may appear lower than expected.

To account for this, two Heckman selection models predicting earnings were run for women, the first selecting on economic activity (being employed or unemployed) and the second selecting on full-time employment. Additional variables included in the selection model were number of children, marital status, and whether their health limited their daily activities (as disability or poor health was another primary reason people were not in the labour market). The mills Lamda's in all regressions are statistically significant, but particularly strong for the US sample. The Lamda's are negative, suggesting that regressions that do not account for selection would have given a downward estimate of the relationship between parents education and income in adulthood for women.

Table 7.7: Heckman two-stage regression predicting log earnings by family background, selecting on probability of being economically active.

	UK			US		
	Model 1	Model 2	Model3	Model 1	Model 2	Model3
Parents have high education	0.231***	0.138***	0.137***	0.176	0.135	0.132
	(0.040)	(0.040)	(0.040)	(0.202)	(0.206)	(0.205)
Degree		0.486***			0.219	
		(0.044)			(0.238)	
Subject studied						
No degree						
STEM			0.565***			0.153
			(0.078)			(0.341)
LEM			0.706***			0.266
			(0.085)			(0.385)
OSSAH			0.391***			0.232
			(0.051)			(0.310)
Ethnicity						
BME	0.370***	0.309***	0.301***			
	(0.110)	(0.106)	(0.105)			
Hispanic				0.445	0.435	0.437
				(0.280)	(0.280)	(0.280)
Black				0.115	0.0891	0.0905
				(0.218)	(0.216)	(0.218)
Cognitive tests	0.147***	0.0893***	0.0894***	0.474***	0.436***	0.437***
	(0.020)	(0.020)	(0.020)	(0.118)	(0.126)	(0.126)
Constant	9.870***	9.781***	9.780***	16.48	16.73	16.97
	(0.036)	(0.036)	(0.036)	(33.429)	(33.323)	(33.263)
Selection model: economic activity						
Parents have high education	-0.00853	-0.0258	-0.0257	-0.0730	-0.0833	-0.0846
	(0.056)	(0.057)	(0.057)	(0.064)	(0.065)	(0.065)
Degree		0.0869			0.0725	
		(0.062)			(0.074)	
Subject studied						
No degree						
STEM			0.223*			0.0305
			(0.122)			(0.115)
LEM			-0.0263			0.0869
			(0.119)			(0.117)
OSSAH			0.0795			0.0931
			(0.074)			(0.103)
Ethnicity						
BME	-0.0625	-0.0715	-0.0727			
	(0.143)	(0.143)	(0.143)			
Hispanic				0.0105	0.00717	0.00782
				(0.084)	(0.085)	(0.085)
Black				0.111*	0.103	0.104
				(0.067)	(0.068)	(0.068)
Cognitive tests	0.111***	0.102***	0.101***	0.0996***	0.0870**	0.0875**
	(0.025)	(0.026)	(0.026)	(0.032)	(0.036)	(0.036)
Children						
1	-0.264***	-0.260***	-0.258***	-0.0641	-0.0585	-0.0587
	(0.080)	(0.080)	(0.080)	(0.089)	(0.089)	(0.089)
2	-0.379***	-0.374***	-0.373***	-0.204***	-0.200**	-0.199**
	(0.070)	(0.070)	(0.070)	(0.077)	(0.077)	(0.077)
3	-0.709***	-0.702***	-0.701***	-0.290***	-0.284***	-0.283***
	(0.078)	(0.078)	(0.078)	(0.085)	(0.085)	(0.085)
4 or more	-1.164***	-1.155***	-1.156***	-0.429***	-0.421***	-0.420***
	(0.091)	(0.092)	(0.092)	(0.093)	(0.094)	(0.094)
Married	-0.0183	-0.0215	-0.0222	-0.0972*	-0.101*	-0.100*
	(0.047)	(0.047)	(0.047)	(0.052)	(0.053)	(0.053)
Poor/limiting health	-0.641***	-0.639***	-0.639***	-0.826***	-0.824***	-0.824***
	(0.046)	(0.046)	(0.046)	(0.079)	(0.079)	(0.079)
Constant	1.566***	1.551***	1.550***	-0.108	-0.159	-0.124
	(0.067)	(0.068)	(0.068)	(9.436)	(9.436)	(9.421)
Mills Lambda	-0.954***	-0.920***	-0.916***	-3.782***	-3.760***	-3.753***
	(0.100)	(0.098)	(0.098)	(0.614)	(0.613)	(0.614)
		5120			4197	

* p < 0.10, ** p < 0.05, *** p < 0.01

Table 7.8: Heckman two-stage regression predicting log earnings by family background, selecting on probability of being in full time work.

	UK			US		
	Model 1	Model 2	Model3	Model 1	Model 2	Model3
Parents have high education	0.241***	0.182***	0.182***	0.156	0.114	0.124
	(0.040)	(0.040)	(0.040)	(0.178)	(0.182)	(0.182)
Degree		0.346***			0.228	
		(0.043)			(0.203)	
Subject studied						
No degree						
STEM			0.452***			0.356
			(0.078)			(0.316)
LEM			0.480***			0.249
			(0.082)			(0.340)
OSSAH			0.260***			0.134
			(0.050)			(0.274)
Ethnicity						
BME	0.251**	0.212**	0.202*			
	(0.111)	(0.106)	(0.106)			
Hispanic				0.248	0.244	0.239
				(0.237)	(0.236)	(0.238)
Black				0.0301	0.00485	-0.00242
				(0.214)	(0.212)	(0.215)
Cognitive tests	0.138***	0.0944***	0.0936***	0.399***	0.360***	0.357***
	(0.021)	(0.021)	(0.021)	(0.118)	(0.120)	(0.121)
Constant	10.23***	10.15***	10.15***	25.29	25.01	24.85
	(0.048)	(0.050)	(0.049)	(26.502)	(26.515)	(26.382)
Selection model: full time employment						
Parents have high education	-0.0476	-0.0898*	-0.0900*	-0.0915	-0.0940	-0.0964
	(0.045)	(0.047)	(0.047)	(0.059)	(0.061)	(0.061)
Degree		0.212***			0.0166	
		(0.050)			(0.065)	
Subject studied						
No degree						
STEM			0.271***			-0.112
			(0.093)			(0.103)
LEM			0.308***			0.100
			(0.099)			(0.110)
OSSAH			0.162***			0.0528
			(0.060)			(0.085)
Ethnicity						
BME	0.163	0.140	0.136			
	(0.126)	(0.127)	(0.127)			
Hispanic				0.110	0.109	0.110
				(0.076)	(0.076)	(0.076)
Black				0.217***	0.215***	0.217***
				(0.064)	(0.065)	(0.065)
Cognitive tests	0.0821***	0.0577***	0.0576***	0.105***	0.102***	0.103***
	(0.021)	(0.022)	(0.022)	(0.030)	(0.033)	(0.033)
Children						
1	-0.761***	-0.753***	-0.752***	-0.0948	-0.0935	-0.0952
	(0.062)	(0.062)	(0.062)	(0.073)	(0.073)	(0.073)
2	-1.012***	-1.003***	-1.003***	-0.279***	-0.278***	-0.276***
	(0.055)	(0.055)	(0.055)	(0.067)	(0.067)	(0.067)
3	-1.207***	-1.194***	-1.194***	-0.406***	-0.405***	-0.402***
	(0.066)	(0.066)	(0.066)	(0.077)	(0.077)	(0.077)
4 or more	-1.473***	-1.457***	-1.458***	-0.599***	-0.597***	-0.594***
	(0.090)	(0.090)	(0.090)	(0.090)	(0.089)	(0.090)
Married	-0.0625	-0.0693*	-0.0699*	-0.124**	-0.125**	-0.125**
	(0.040)	(0.040)	(0.040)	(0.050)	(0.051)	(0.050)
Poor/limiting health	-0.335***	-0.331***	-0.330***	-0.718***	-0.718***	-0.715***
	(0.043)	(0.043)	(0.043)	(0.080)	(0.080)	(0.080)
Constant	0.819***	0.782***	0.782***	-6.252	-6.265	-6.118
	(0.051)	(0.052)	(0.052)	(8.505)	(8.503)	(8.479)
Mills Lambda	-0.261***	-0.250***	-0.248***	-2.138***	-2.113***	-2.123***
	(0.055)	(0.054)	(0.054)	(0.568)	(0.566)	(0.571)
	5120			4197		

* p < 0.10, ** p < 0.05, *** p < 0.01

Findings from the selection model showing the associates of economic activity or full time work run in the directions we would expect. Women with more children had lower likelihood of being economically active and employed full time, and marital status was associated with lower likelihood of being employed in the US, and lower likelihood of working full time in both countries. Having limiting poor health was strongly associated with being both inactive and not working full-time.

In terms of the substantive findings, overall parent's education was not associated with employment status, except that UK women were less likely to be working full time if their parents were more educated. Most importantly, and in line with the main results, we see no significant differences between models two and three when accounting for selection into economic activity or full-time employment.

The selection models are different from the main models in that the estimates are not weighted, as heckman regression does not allow probability weights. These models likely also lack power compared to the main regression models, because the correction term introduced to account for selection and the variables in the earnings model, will be correlated. This leads to an inflation of the standard errors (Moffitt 1999; Stolzenberg and Relles 1990). Furthermore, we cannot assume that the error terms will be independent, as there will likely be some unobservable variables associated with both employment status and earnings. Thus, given the similarities in substantive findings, the preferred models would be those that do not account for selection.

7.7 Discussion

This study compared stratification in field of study by family background in the UK and the US, using two comparable nationally representative surveys. Whilst no associations between parental education and field of study were found in the UK, women in the US whose parents were less educated were more likely to hold a STEM or LEM degree, rather than and OSSAH degree. These differences were robust to controls for ability measured through early cognitive test scores, and the inclusion of other measures of family background. In the UK, the most disadvantaged women (those who were eligible for Free School Meals) were least likely to study OSSAH subjects.

In chapter five I showed in a UK cohort that parents education *was* associated with subject choice for women, with more advantaged women more likely to study arts and humanities subjects. The results in this chapter run contrary to those in chapter five, possibly due to the different timings of university entry and differences in the policy contexts. Importantly, the expansion of higher education was just beginning in the late 1980's and early 1990s, when this study's cohort were attending university, but was well underway by the time the cohort in chapter five were attending university. There have also been huge policy changes in fees charged to students and financial support available, potentially making subject choices more pertinent for later cohorts. In many ways, the university system in the UK now is more similar to the US system (with high fees and high levels of attendance) than when the BCS70 cohort would have attended university.

For men in both countries there was little association between social background and field of study, suggesting family background only influenced young women's choices. This is in line with results from chapter five, which suggested that the relationship between social background and subject chosen was stronger for women than men. One measure of family background that did predict choices of US men was family type. Results suggest that growing up without a father has a negative association with propensity to study STEM, but, possibly through a role model effect, has a positive association with women's propensity to study more lucrative careers. It is hypothesised the negative relationship between STEM and not having a father may be driven by the lack of 'STEM capital' within in family (see Archer, Dawson, DeWitt, Seakins, & Wong, 2015). Because men are more likely to work in, or have a strong interest in STEM, the lack of a father in the household may mean children are less able to explore any interest in STEM outside of school. Further research, particularly looking specifically at parent's occupational field, could help to further understand the mechanisms behind this relationship.

This study also set out to test the extent that stratification by field of study explain differences in earnings. Both men and women with less educated parents earned less than their more advantaged counterparts, and whilst degree attainment reduced this gap (or eliminated it for US women), taking account of field of study did not further explain disparities. This is particularly surprising for US women, who were more likely to choose more lucrative subjects if their parents were less educated, and thus should be further closing the earnings gap. It is likely, however, that returns to the same degrees differ by family background, and that more advantaged women can draw on networks and social or financial capital to help them succeed even with less lucrative majors.

The extent that country features explain differences in associations cannot be formally tested, but results run in line with predictions. Firstly, in the US, where students are not stratified into subjects at an early age, less advantaged women are more likely to study STEM subjects. In the UK, women from disadvantaged background may have 'closed the door' to STEM at a much earlier age, due to stereotypes about both the suitability of women in STEM (Beasley & Fischer, 2012; Shapiro & Williams, 2012), and of relative ability required to study STEM subjects (Archer et al., 2013). Thus, only in the US are women able to fully explore all options before choosing a subject. Secondly, there was also likely more concern about returns after university in the US due to the high cost of study, and this may be particularly important for people from less advantaged backgrounds. Finally, the larger proportion of people in the US studying any degree would likely strengthen associations between background and choices, as students would have greater impetus to differentiate themselves along other criteria.

Results do not support the *effectively maintain inequality* hypothesis; more advantaged individuals did not choose subjects which would give them better access to high income careers, but instead support a *safety net* hypothesis for US women, suggesting they are less likely to be concerned by later outcomes in making course choices. It remains possible that the *effectively maintained inequality* hypothesis holds for horizontal stratification along other domains, for example status of university, and access to these may even mitigate the importance of field of study. For example, knowledge that they can access high income careers after studying any subject at a highly prestigious university, or that they will be able to pursue postgraduate study in a lucrative subject, may have caused advantaged people to be less concerned about undergraduate field of study in trying to maintain their advantage. As yet, there remains a debate in the literature over whether social disparities in admittance to selective universities is entirely driven by prior achievement (Anders, 2012; Boliver, 2013; Chowdry, Crawford, Dearden, Goodman, & Vignoles, 2013; Jerrim, Chmielewski, & Parker, 2015; Jerrim & Vignoles, 2015), however further research in this area would help fully understand the ways in which advantaged families maintain their social position in the context of educational expansion.

It is also possible that associations may have changed dramatically over time, and that people in these surveys attended university too early for students to start differentiating themselves. The intake of students in both countries has increased dramatically. However, by its nature this question can only be addressed with a suitable time lag that allows graduates to establish their place in the earnings distribution. Whilst there is not, currently, suitable survey data that allows such analysis, the availability of linked education and labour market data would make this analysis possible. The availability of linked administrative and survey data would greatly enhance researchers ability to answer these questions, as would any attempts to model expected lifetime earnings based on career information from younger graduates. For example the nature of their employment and whether their early career choices have strong potential for career progression and higher earnings in the future, regardless of current earnings.

Chapter 8: Conclusions

8.1 Overview and summary

This thesis has contributed to the literature on educational inequalities, field of study and the intergenerational transmission of earnings advantage in a number of ways.

Firstly, in chapter three, myself and my co-author Rose Cook contribute to methodological work in quantitative research into educational inequalities by reviewing articles applying the concept of intersectionality in their work, giving an overview of the contributions and limitations of this work, and making recommendations for further use of the concept. In this joint-authored chapter, I argue that researchers should more often consider the intersectional relationships driving inequalities, and more readily acknowledge that a persons 'set' of characteristics, as opposed to each characteristic in isolation, are uniquely associated with choices and outcomes.

I also make a number of recommendations for researchers, including to be more explicit about when the concept is used; for academics, research councils and policy makers to work together to facilitate the linkages between survey and administrative data so that researchers can analyse more detailed axes of social inequality; and that researchers make use of quantitative data to identify *drivers* of inequalities, by exploring how associations change depending on social context. This chapter should not only help inform academics' analysis of inequalities and offer motivation for the study of intersectional patterns, but also act as further argument for policy makers and data controllers to make the data needed for this analysis more available.

The three empirical chapters in this thesis are all concerned with the relationship between social background and field of study in post compulsory education, with each exploring a different aspect of this relationship.

The first empirical paper, chapter five, draws on the concept of intersectionality outlined in chapter three to analyse the associations between student characteristics and field of study at A level and university, using Next Steps data. I find that white students are less likely to study at least one STEM subject at A level than students from all other ethnic groups. Whilst other studies have identified that Asian students are more likely to study STEM, I find that both black Caribbean and African students are also more likely to choose STEM once controlling for prior attainment. Fathers' educational attainment was positively associated, and Mothers' educational attainment negatively associated, with STEM study at A level. Similar ethnic patterns were found for degree subject choice, with white students most likely to study arts and humanities subjects. Mothers' educational attainment was also associated with choice of arts and humanities degrees.

I then considered whether there were interactions between gender, social background and ethnicity. Women from more advantaged backgrounds were even less likely to study SLB subjects, and more likely to study STEM, compared with men from similar backgrounds. This suggests that the gender differences in subjects studied were smaller for more advantaged students. Indian students from more privileged backgrounds were less likely to choose arts and humanities over STEM, and black African women were particularly less likely to choose arts and humanities. I discuss the possible reasons for these specific interactions, including that student's family and home environment differentially influence gender attitudes, or students' feelings of personal control over their futures.

Chapter six extends this analysis by unpicking the mechanisms behind differences in subject choices by social background. Whilst there is a very large literature exploring the reasons women are less likely to study science and maths, the finding that there is also a social background gradient to subject choices is relatively new, and consequently little research has considered the reasons for these differences. Informed by the psychological research into gender disparities in choices, I considered whether personal preferences and beliefs could go some way to explaining disparities by levels of advantage. In contrast to findings for gender disparities, I do not find evidence that personal attitudes help to explain differences in choices. I do find, however, that students were more likely to choose subjects based on preferences if they were from more advantaged backgrounds. This finding was robust to controls for prior attainment, previous qualification type and other characteristics. This suggests that the processes driving disparities by social background are different to those driving gender disparities, and therefore interventions aimed at improving uptake amongst under-represented groups will likely need to employ different approaches to be successful.

Chapter seven explored another aspect of the relationship between subject studied and social background; the extent that differences in choices explain the intergenerational transmission of advantage from parents to children. In other words, if there were no differences in subject choice by background, would the relationship between parents' education and children's earnings be reduced? In exploring this question, I first compared the relationship between social background and subject choice at university in the UK and the US. A stronger association was identified in the US; young women from less advantaged backgrounds were more likely to choose both more lucrative (LEM) and typically male dominated (STEM) subjects compared to arts and humanities. In the UK, only the most disadvantaged women (those who were eligible for free school meals) were more likely to study STEM. I suggest this may be due to the increased importance of subject studied for less advantaged women, who lacked the parental and family resources to achieve in whatever subject they studied. This ties in with findings from chapter six, where more advantaged young people are more likely to choose subjects they enjoy.

I also find that controlling for subject studied, over just degree attainment, did not help to further explain the association between parents' education and children's earnings. This was particularly surprising in the case of US women; disadvantaged women in the US were choosing subjects that would be typically associated with higher returns, therefore it was expected that the gap would increase once controlling for subject choice. It is of course possible that there are heterogeneous returns to subjects depending on background, and students with strong family connections and support may do well even if their subject area is typically associated with lower returns.

Several themes emerged from the three empirical papers. In all chapters, a relationship between family background and subject choices at university was observed. Whilst chapter five set out to test the relationship between a broad range of indicators of family background, parents' education had the strongest relationship with young people's choices, allowing a more parsimonious analysis in chapters six and seven. Focusing on the most disadvantaged students would likely have produced different result, particularly when controlling for academic attainment. A question that remains from all papers is how far initial 'filtering' of young people out of education has impacted results. The sample of students attending university is already very different in terms of social background than the population (as outlined in chapter four). Using the National Pupil Database (NPD), Strand (2017) finds that differences in attainment in science at age 18-19 between students who had ever been entitled to Free School Meals (FSMs), compared to those who had not, are reduced when accounting for overall participation. This filtering effect was more pronounced for the BCS70 sample of students, where fewer young people attended university, than for the Next Steps sample. Whilst this can offer insights into why results for these two cohorts were so different, it also suggests attention needs to be paid to earlier choices and streaming of students.

Chapter seven incorporates an indicator for whether people had ever claimed FSM, along with focusing on subject choices of a different cohort of young people, and across two countries. Whilst the relationships are still observed in these cohorts, the direction of relationships and the variables associated with choices differed. This raises an important point about the timing of analysis, and the applicability of historical findings to current cohorts of students. The university system changed substantially in the time between the two cohort studies, and has undergone large changes in the years after attendance of the Next Steps sample at university. Whilst the results are very much relevant in understanding the choices of young people now at the beginning of their careers, it remains unclear whether the same insights can be applied to students choosing their subjects now.

Chapters five and seven both employ the principle of intersectionality to analysis, as recommended in chapter three. There were strong theoretical and empirical motivations to test for interactions between young people's characteristics in determining choices. Both papers find an interaction between gender and social background in predicting subject choices, albeit in different directions. However, few interactions were found between ethnicity and subject choices in chapter five, despite much research suggesting a strong interaction between these two characteristics in predicting academic attainment. Despite this null result, it remains an important finding that we do not necessarily need to consider ethnicity as a separate factor when thinking about the ways young people's social background may influence choices.

8.2 Future research

8.2.1 Validity and timeliness of findings

The results from this thesis primarily focus on a very recent cohort, Next Steps, who were born between 1989 and 1990. The subject choices of these young people will have effects on society for many years, through their participation in the labour market and their occupational pathways, so it's remains important to understand which characteristics influenced subject choices. However, it is unclear whether results could be applicable to young people entering university now. Particularly given policy changes around university funding, including the increase in student fees in 2012 and the abolition of maintenance grants (payments offered to students from low income families attending university with no expectations of repayment) in 2016, students' reasoning behind subject choices may have changed. Students from very low-income backgrounds will now have a much larger debt burden then their more advantaged peers following the conversion of grants to loans, which may influence the extent that their subjects choices are impacted by financial considerations. These considerations could include the feasibility of combining study and work (some courses, particularly STEM subjects, having higher contact hours), about the difficultly of courses or likelihood of achieving high grades, and about predicted economic opportunities after graduation.

Overall, the question of the extent that stratification into subjects influences later inequalities could be explored with more contemporary data. Higher education data has recently been linked with tax records held by Her Majesty's Revenue and Customs (HMRC) and the Department of Work and Pensions (DWP) to create the Longitudinal Education Outcomes (LEO) dataset. This would allow comprehensive assessment of the interplay between income dynamics, individual's complete educational career, and their social background. Next Steps has also released data on outcomes for participants at age 25, making it possible to measure the early effects of subject choices on inequalities. Both datasets have their drawbacks. The family background characteristics available in the LEO datasets are not as comprehensive as those measured in survey datasets. Whilst Next Steps does offer a very broad view of young people's background, occupational outcomes are reported very early in their career, and it would be expected that inequalities would be very small at the start of individuals' careers (a pattern seen in the results using BCS70 data outlined in chapter seven). Thus, any effect of subject studied on outcomes would likely be understated. Furthermore, occupational outcomes in Next Steps are self-reported, and would not be as accurate as administrative records (as is also an issue with the BCS70 dataset). An ideal scenario would be for both the data sources to be used and compared to quantitatively assess the implications of the weaknesses of each.

The Longitudinal Studies Strategic Review Report (2017), conducted by an independent and international panel, recommends that the Economic and Social Research Council develop and maintain a 'data-spine,' which will hold unique identifiers for individuals covering as much of the population of the UK as possible. This resource would then be used to more easily link administrative datasets with one another and with survey data. This follows the model of other countries, including Canada, Australia, and New Zealand, whilst also allowing better use of the UK's unique survey resources. If these recommendations are implemented, this will both increase the breadth of research questions that can be addressed quantitatively and the accuracy of results. Furthermore, it would allow researchers to study smaller groups of people within the population, and apply intersectionality as a method more readily (as discussed in chapter three).

There is scope for future research focusing on heterogeneous returns to subjects, answering the question of which subjects offer similar returns whatever your background, and which subjects have very different returns for advantaged and disadvantaged people. As outlined by analysts at the Department of Education using LEO data, some subjects have relatively similar returns for all graduates (i.e. medicine), whilst for other subjects there is a large difference between the highest and lowest earning students. Economics and law are associated with higher, but also very heterogeneous returns, and future research could consider whether it is the less advantaged students who are less likely to attain the highest paying jobs, even after studying lucrative subjects (Department for Education, 2017). Some subject specific research has looked into this, and rhetoric is particularly focused on differential opportunity to succeed in the arts and humanities, even with similar qualifications, by social background (e.g. Friedman, Brien, & Laurison, 2016).

8.2.2 Exploration of the mechanisms driving results

I have identified what *does not* explain differences in subject choices by social background in this thesis, but not what *does.* Whilst null results are important in ruling out hypotheses, and dispelling common assumptions, this remains a large gap in the literature. I have suggested a number of reasons the gap remains, including differences in parenting, or exposure to science by background. Future research could attempt to empirically measure this, by for example asking parents how much time they spend on extra curricula activities in different subject areas. In the US, Ma (2009) found that parental involvement in subject specific domains was associated with choices, but this did not explain why less advantaged young people were more likely to choose majors associated with higher expected earnings. To my knowledge, similar analysis has not been conducted in the UK context. Furthermore, a quantitative measure of the concept of 'science capital' could be constructed, to tests whether this explains disparities, as suggested in some qualitative work (e.g. Archer, Dawson, DeWitt, Seakins, & Wong, 2015). This could include information not only on extra curricula activities, but also parents' own field of education and interests. Longitudinal surveys could include questions not only on parents' level of education, but also their field, to facilitate this analysis.

8.3 Policy implications

Findings suggest that the social gradient in uptake of STEM subjects is stronger for women than men. Currently, the majority of interventions aimed at getting more women into STEM are targeted either at all girls and young women, those who are particularly high achieving (and thus likely from more advantaged backgrounds). Findings from this research suggest that more focus needs to be directed towards interventions targeting less advantaged girls. Students from lower SES backgrounds are less likely to receive good quality careers advice, but are the ones that need it most (Archer, et al. 2013). A clear way to get more students to study STEM would be to give them knowledge of the wide range of careers available upon graduation after studying STEM, or of the subjects they need to study at GCSE and

A-level to study STEM at university. Advantaged students, particularly those whose parents work in STEM spheres, are most likely to already have access to this knowledge.

Overall, all young people could benefit from more information about the returns and career opportunities associated with different subjects. However, these returns differ depending on students' backgrounds, with more advantaged young people more likely to gain well-paid employment whatever they choose to study. The analysis in this thesis suggests that students are already likely somewhat aware of this, choosing subjects with higher returns, avoiding arts and humanities, and placing less importance on the potential intrinsic returns of study. This inequality will only be resolved by focusing more broadly on inequalities in the labour market. Differences in motivations for study will be unnecessary if young people were afforded similar opportunities regardless of social background, gender, and ethnicity.

Bibliography

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123-129.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica, 65*(5), 1005-1027.

Allison, P. D. (2001). *Missing data (Quantitative applications in the social sciences).* London UK: Sage Publications.

Alon, S. (2007). Overlapping disadvantages and the racial/ethnic graduation gap among students attending selective institutions. *Social Science Research*, *36*(4), 1475-1499.

Altonji, J. G., Kahn, L. B., & Speer, J. D. (2016). Cashier or consultant? Entry labor market conditions, field of study, and career success. *Journal of Labor Economics*, *34*(1), 361-401.

Anders, J. (2012). The link between household income, university applications and university attendance. *Fiscal Studies*, *33*(2), 185–210.

Anders, J. (2017). The influence of socioeconomic status on changes in young people's expectations of applying to university. *Oxford Review of Education*, *43*(4), 381–401.

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, 44*(1), 75-93

Andrews, J., Robinson, D., & Hutchinson, J. (2017). *Closing the gap? Trends in educational attainment and disadvantage.* London: Education Policy Unit.

Antill, J. K., Cunningham, J. D. & Cotton, S. (2003). Gender-role attitudes in middle childhood: In what ways do parents influence their children? *Australian Journal of Psychology, 55*(3), 148–153.

Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). "Science capital": A conceptual, methodological, and empirical argument for extending bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching*, *52*(7), 922–948.

Archer, L., DeWitt, J., & Dillon, J. (2014). 'It didn't really change my opinion': exploring what works, what doesn't and why in a school science, technology, engineering and mathematics careers intervention. *Research in Science & Technological Education, 32*(1), 35-55.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus how families shape children's engagement and identification with science. *American Educational Research Journal, 49*(5), 881-908.

Archer, L., & Francis, B. (2007). *Understanding minority ethnic achievement: race, gender, class and "success."* London: Routledge.

Archer, L., Francis, B., Moote, J., DeWitt, J., & Yeomans, L. (2016). The "exceptional" physics/engineering girl: A sociological analysis of longitudinal data from girls aged 10–16 to explore gendered patterns of post-16 participation. *American Educational Research Journal, 54*(1), 88 - 126.

Archer, L., & Moote, J. (2016). *ASPIRES 2 project spotlight: year 11 students' views of careers education and work experience.* London: King's College.

Archer, L., Osborne, J., DeWitt, J., Dillon, J., Wong, B., & Willis, B. (2013). *ASPIRES. Young people's science and career aspirations, age 10 – 14.* Retrieved from: *https://www.kcl.ac.uk/sspp/departments/education/research/aspires/ASPIRES-final-report-December-2013.pdf*

Barone, C., & van de Werfhorst, H. G. (2011). Education, cognitive skills and earnings in comparative perspective. *International Sociology*, *26*(4), 483–502.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*(4), 427–448.

Bécares, L., & Priest, N. (2015). Understanding the influence of race/ethnicity, gender, and class on inequalities in academic and non-academic outcomes among eighth-grade students: findings from an intersectionality approach. *PloS one*, *10*(10), e0141363.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., van der Erve, L., Sibieta, L., Vignoles, A., Walker, I., & Zhu, Y. (2018). *The relative labour market returns to different degrees.* London: Institute for Fiscal Studies (IFS).

Bernardi, F., & Ballarino, G. (2016). *Education, occupation and social origin: a comparative analysis of the transmission of socio-economic inequalities*. Cheltenham: Edward Elgar Publishing.

Berrington, A., Roberts, S., & Tammes, P. (2016). Educational aspirations among UK young teenagers: Exploring the role of gender, class and ethnicity. *British Educational Research Journal*, *42*(5), 729-755.

Bhattacharyya, G., Ison, L. & Blair, M. (2003). *Minority ethnic attainment and participation in education and training: The evidence.* Nottingham:

Department for Education and Skills.

Blanden, J., & Gregg, P. (2004). Family income and educational attainment: a review of approaches and evidence for Britain. *Oxford Review of Economic Policy, 20*(2), 245-263.

Blanden, J., Gregg, P., & Machin, S. (2005). *Intergenerational mobility in europe and north america.* London: London School of Economics, Centre for Economic Performance.

Blanden, J., & Macmillan, L. (2016). Educational inequality, educational expansion and intergenerational mobility. *Journal of Social Policy*, *45*(4), 589–614.

Blundell, R., Dearden, L., Goodman, A., & Reed, H. (2000). The returns to higher education in britain: evidence from a british cohort. *The Economic Journal*, *110*(461), 82–99.

Boaler, J., Altendorff, L. & Kent, G. (2011). Mathematics and science inequalities in the United Kingdom: When elitism, sexism and culture collide. *Oxford Review of Education, 37*(4), 457–484.

Boalt, G., & Janson, C. G. (1953). A selected bibliography of the literature on social stratification and social mobility in sweden. *Current Sociology*, *2*, 306–327.

Bøe M.V., Henriksen E.K. (2015). Expectancy-value perspectives on choice of science and technology education in late-modern societies. In: Henriksen E., Dillon J., Ryder J. (Ed.) *Understanding student participation and choice in science and technology education* (pp 17-29). Dordrecht: Springer.

Boliver, V. (2013). How fair is access to more prestigious UK universities? *British Journal of Sociology*, *64*(2), 344–364.

Boliver, V. (2015). Are there distinctive clusters of higher and lower status universities in the UK? *Oxford Review of Education*, *41*(5), 608–627.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*(1), 605–634.

Bostwick, V. (2016). Signaling in higher education: The effect of access to elite colleges on choice of major. *Economic Inquiry*, *54*(3), 1383–1401.

Botcherby, S., & Buckner, L. (2012). *Women in Science, Technology, Engineering and Mathematics: from Classroom to Boardroom.* Bradford: WISE Campaign.

Botelho, A., & Pinto, L. C. (2004). Students' expectations of the economic returns to college education: results of a controlled experiment. *Economics of Education Review, 23*(6), 645-653.

Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in Western society.* New York: Wiley.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Cambridge, Massachusetts: Harvard University Press.

Bradley, K., & Charles, M. (2003). Uneven inroads: understanding women's status in higher education. *Research in the Sociology of Education*, 247-274.

Bratti, M., Naylor, R., & Smith, J. (2008). Heterogeneities in the returns to degrees: evidence from the british cohort study 1970. *DEAS, University of Milan, Departmental Working Paper No. 2008-40*.

Breen, R. (2010). Educational expansion and social mobility in the 20th century. *Social Forces*, *89*(2), 365-388.

Breen, R., & Goldthorpe, J. H. (1997). Explaining educational differentials: towards a formal rational action theory. *Rationality and Society*, *9*(3), 275–305.

Breen, R., & Jonsson, J. O. (2005). Inequality of opportunity in comparative perspective: recent research on educational attainment and social mobility. *Annual Review of Sociology*, *31*, 223–243.

Breen, R., Luijkx, R., Müller, W., & Pollak, R. (2009). Nonpersistent inequality in educational attainment: evidence from eight european countries. *American Journal of Sociology*, *114*(5), 1475–1521.

Breen, R., Luijkx, R., Müller, W., & Pollak, R. (2010). Long-term trends in educational inequality in europe: Class inequalities and gender differences. *European Sociological Review*, *26*(1), 31–48.

Breen, R., & Yaish, M. (2006). Testing the Breen-Goldthorpe model of educational decision making. In: Morgan SL, Grusky D. B. and Fields G. S. (Ed.) *Mobility and inequality: frontiers of research from sociology and economics* (pp 232-58). Stanford, CA: Stanford University Press.

Britton, J., Dearden, L., Shephard, N., & Vignoles, A. (2016). How english domiciled graduate earnings vary with gender , institution attended , subject and socio-economic background*. IFS Working Paper W16/06*

Broughton, N. (2013). *In the balance: The STEM human capital crunch.* Retrieved from: http://www.smf.co.uk/wp-content/uploads/2013/03/Publication-In-The-Balance-The-STEM-human-capital-crunch.pdf

Browne, I., & Misra, J. (2003). The intersection of gender and race in the labor market. *Annual Review of Sociology*, *29*, 487–513.

Buchmann, C., DiPrete, T., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology, 34,* 319-337.

Bukodi, E., Erikson, R., & Goldthorpe, J. H. (2014). The effects of social origins and cognitive ability in educational attainment: evidence from britain and sweden. *Acta Sociologica*, *57*(4), 293–310.

Bukodi, E & Goldthorpe, J. H. (2013). Decomposing social origins: the effects of parents class, status, and education on the educational attainment of their children. *European Sociological Review 29*(5): 1024–1039.

Bukodi, E., Goldthorpe, J. H., Waller, L., & Kuha, J. (2015). The mobility problem in Britain: new findings from the analysis of birth cohort data. *The British journal of sociology, 66*(1), 93-117.

Burgess, S., Wilson, D., & Worth, J. (2009). *Passing through school: the evolution of attainment of England's ethnic minorities.* Bristol: CMPO.

Campbell, T. (2015). Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment. *Journal of Social Policy*, *44*(3), 517–547.

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics, 3,* 1801–1863.

CaSE (2014). *Improving diversity in STEM.* Retrieved from: http://www.sciencecampaign.org.uk/resource/ImprovingDiversityinSTEM2014.html

Catsambis, S. (1994). The path to math: Gender and racial-ethnic differences in mathematics participation from middle school to high school. *Sociology of Education*, *67*(3), 199-215.

CBI. (2014). *Engineering our future: stepping up the urgency on stem*. London: Confederation for British Industry.

Cech, E., Rubineau, B., Silbey, S., & Seron, C. (2011). Professional role confidence and gendered persistence in engineering. *American Sociological Review*, *76*(5), 641–666.

Charles, M. (2017). Venus, Mars, and math: Gender, societal affluence, and eighth graders' aspirations for STEM. *Socius,* 3, 1-16.

Charles, M., & Bradley, K. (2002). Equal but separate? A cross-national study of sex segregation in higher education. *American Sociological Review*, *67*(4), 573–599.

Charles, M., and Bradley, K. (2004). Uneven inroads: Understanding women's status in higher education. *Research in Sociology of Education, 14*, 247-274.

Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, *114*(4), 924–976.

Checchi, D., & van de Werfhorst, H. G. (2017). Policies, skills and earnings: how educational inequality affects earnings inequality. *Socio-Economic Review, 16* (1), 137-160.

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2018). *Race and economic opportunity in the united states: an intergenerational perspective* (Working paper no. 24441). Cambridge, MA: National Bureau of Economic Research.

Chevalier, A. (2011). Subject choice and earnings of UK graduates. *Economics of Education Review*, *30*(6), 1187–1201.

Chipuer, H. M., Rovine, M. J., & Plomin, R. (1990). LISREL modeling: Genetic and environmental influences on IQ revisited. *Intelligence*, *14*(1), 11–29.

Choo, H. Y., & Ferree, M. M. (2010). Practicing intersectionality in sociological research: A critical analysis of inclusions, interactions, and institutions in the study of inequalities. *Sociological Theory*, *28*(2), 129–149.

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. Series A: Statistics in Society*, *176*(2), 431–457.

Chowdry, H., Crawford, C. & Goodman, A. (2011). The role of attitudes and behaviours in explaining socio-economic differences in attainment at age 16. *Longitudinal and Life Course Studies, 2*(1), 59–76.

Codiroli, N. (2015). *Inequalities in students' choice of stem subjects.* CLS Working Paper 2015/6, Centre for Longitudinal Studies, UCL Institute of Education.

Codiroli Mcmaster, N. (2017). Who studies STEM subjects at A level and degree in England? An investigation into the intersections between students' family background, gender and ethnicity in determining choice. *British Educational Research Journal*, *43*(3), 528–553.

Codiroli Mcmaster, N. (2017a, July 10). Women are less likely to study STEM subjects – but disadvantaged women are even less so. [Blog post]. Retrived from: http://blogs.lse.ac.uk/politicsandpolicy/who-studies-stem/

Codiroli Mcmaster, N. (2017b, August 8). Which subjects bring the best career outcomes for UK university students? [Blog post]. Retrived from: http://blogs.lse.ac.uk/businessreview/2017/08/08/which-subjects-bring-the-best-career-outcomes-for-uk-university-students/

Codiroli Mcmaster, N., & Cook, R. (2018). The contribution of intersectionality to quantitative research into educational inequalities. *Review of Education.* (Online First). doi: 10.1002/rev3.3116

Coe, R., Searle, J., Barmby, P., Jones, K., & Higgins, S. (2008). *Relative difficulty of examinations in different subjects.* Durham, UK: Centre for Educational Management.

Collings, J. & Smithers, A. (1984). Person orientation and science choice*. European Journal for Science Education, 6*, 55–65.

Conti, G., Heckman, J., & Urzua, S. (2010). The education-health gradient. *American Economic Review*, *100*(2), 234–238.

Crawford, C. (2014). Socio-economic differences in university outcomes in the UK: drop-out, degree completion and degree class. *Institute for Fiscal Studies*, (W14/31).

Crawford, C., & Greaves, E. (2015). *Socio-economic, ethnic and gender differences in HE participation.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/474273/BIS-15-85-socio-economic-ethnic-and-gender-differences.pdf

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum, 1989*(1), 139-167.

Crenshaw, K. (1991). Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, *43*(6), 1241-1299.

Crompton, R., & Lyonette, C. (2005). The new gender essentialism - Domestic and family "choices" and their relation to attitudes. *British Journal of Sociology, 56(4), 601-620*.

Cunha, F., & Heckman, J. J. (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association*, *7*(2–3), 320–364.

Davies, P., Mangan, J., Hughes, A., & Slack, K. (2013). Labour market motivation and undergraduates' choice of degree subject. *British Educational Research Journal, 39*(2), 361-382.

Davies, S., & Guppy, N. (1997). Fields of study, college selectivity, and student inequalities in higher education. *Social Forces*, *75*(4), 1417–1438.

Davis, K. (2008). Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful. *Feminist Theory, 9*(1), 67–85.

Davis, S. N., & Greenstein, T. N. (2004). Cross-national variations in the division of household labor. *Journal of Marriage and Family*, *66*(5), 1260–1271.

Dekkers, H. P. J. M., Bosker, R. J., & Driessen, G. W. J. M. (2000). Complex inequalities of educational opportunities: A large-scale longitudinal study on the relation between gender, social class, ethnicity and school success. *Educational Research and Evaluation*, *6*(1), 69-82.

Department for Business, Innovation and Skills. (2015). *2010 to 2015 government policy: public understanidng of science and engineering.* London: The Stationary Office

Department for Education. (2011). *LSYPE User Guide to the Datasets: Wave 1 to Wave 7.* London: The Stationary Office

Department for Education (2014). *Launch of new high-quality post-16 maths qualifications.* Retrieved from: https://www.gov.uk/government/news/launch-of-new-high-quality-post-16-maths-qualifications

Department for Education (2016). *The lives of young carers in England.* Retrieved from: https://www.gov.uk/government/publications/the-lives-of-young-carers-in-england

Department for Education. (2017). *Employment and earnings outcomes of higher education graduates by subject and institution: experimental statistics using the Longitudinal Education Outcomes (LEO) data*. London: The Stationary Office.

DeWitt, J., Archer, L., & Osborne, J. (2013). Nerdy, brainy and normal: Children's and parents' constructions of those who are highly engaged with science. *Research in Science Education, 43*(4), 1455-1476.

DeWitt, J., Osborne, J., Archer, L., Dillon, J., Willis, B., & Wong, B. (2013). Young children's aspirations in science: The unequivocal, the uncertain and the unthinkable. *International Journal of Science Education, 35*(6), 1037-1063.

Dickerson, A., & Popli, G. (2016). Persistent poverty and children's cognitive development: Evidence from the UK Millennium Cohort Study. *Statistics in Society: Series A*, *179*(2), 535–558.

Dilnot, C. (2016). How does the choice of A-level subjects vary with students' socio-economic status in English state schools? *British Educational Research Journal, 42*(6), 1081-1106.

DiPrete, T. A., Erikson, R., & Goldthorpe, J. H. (1993). The constant flux: a study of class mobility in industrial societies. *Contemporary Sociology*, *22*(4), 536.

Dolton, P. J. & Vignoles, A. (2002). The return on post-compulsory school mathematics study. *Economica, 69*(273), 113–142.

Dynarski, S. M. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *American Economic Review*, *93*(1), 279-288.

Eccles, J. (1983). *Expectancies, values, and academic behaviors.* In J. T. Spence (Ed.), Achievement and achievement motives: Psychological and sociological approaches (pp. 75-146). San Francisco, California: W. H. Freeman.

Eccles, J. (1983). Female achievement patterns: Attributions, expectancies, values, and choice. *Journal of Social Issues, 1*, 1-22.

Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices*. International Journal of Behavioral Development, 35*(3), 195-201.

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*(1), 109-132.

Economic and Social Research Council. (2017). *ESRC Longitudinal Studies Review 2017 Specification*. Retrieved from: https://esrc.ukri.org/news-events-and-publications/publications/corporate-publications/longitudinal-studies-review-2017

Else-Quest, N. M., & Hyde, J. S. (2016). Intersectionality in quantitative psychological research: I. Theoretical and epistemological issues. *Psychology of Women Quarterly*, *40*(2), 155–170.

England, P., Allison, P., Li, S., Mark, N., Thompson, J., Budig, M. J., & Sun, H. (2007). Why are some academic fields tipping toward female? The sex composition of u.s. Fields of doctoral degree receipt, 1971-2002. *Sociology of Education*, *80*(1), 23–42.

England, P., Farkas, G., Kilbourne, B. S., & Dou, T. (1988). Explaining occupational sex segregation and wages: findings from a model with fixed effects. *American Sociological Review*, *53*(4), 544.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2007). Early schooling: The handicap of being poor and male. *Sociology of Education*, *80*(2), 114-138.

Equality Challenge Unit. (2014). *Equality in higher education: Statistical report 2014.* London: Equality Challenge Unit.

Equality Challenge Unit. (2016). *Equality in higher education: Statistical Report 2016.* London: Equality Challenge Unit.

Farré, L., & Vella, F. (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica*, *80*(318), 219–247.

Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, *70*(277), 73–97.

Ferree, M. M., & Hall, E. J. (1996). Rethinking stratification from a feminist perspective: Gender, race, and class in mainstream textbooks. *American Sociological Review*, *61*(6), 929-950.

Few-Demo, A. L. (2014). Intersectionality as the "new" critical approach in feminist family studies: Evolving racial/ethnic feminisms and critical race theories. *Journal of Family Theory & Review*, *6*(2), 169–183.

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school Context. *Journal of Research on Adolescence*, *20*(2), 507–537.

Friedman, S., Brien, D. O., & Laurison, D. (2016). "Like skydiving without a parachute ": how class origin shapes occupational trajectories in british acting. *Sociology, 51*(5), 992-1010.

Friedman, S., Laurison, D., & Macmillan, L. (2017). *Social mobility , the class pay gap and intergenerational worklessness: new insights from the labour force survey*. London: Social Mobility Commission.

Frost, S., Reiss, M. & Frost, J. (2005). Count me in! Gender and ethnic minority attainment in school science. *School Science Review, 86*(316), 105–112.

Gayle, V., Berridge, D. & Davies, R. B. (2003). *Econometric analysis of the demand for higher education.* London: Department for Education and Skills.

Gerber, T. P., & Cheung, S. Y. (2008). Horizontal stratification in postsecondary education: forms, explanations, and implications. *Annual Review of Sociology*, *34*(1), 299–318.

Gilbert, J. (2006). *Science education in schools: Issues, evidence and proposals.* Retrieved from: https://www.stem.org.uk/resources/elibrary/resource/27208/science-education-schools-issues-evidence-and-proposals

Gillborn, D. (2015). Intersectionality, critical race theory, and the primacy of racism. *Qualitative Inquiry*, *21*(3), 277–287.

Gillborn, D., Demack, S., Rollock, N., & Warmington, P. (2017). Moving the goalposts: Education policy and 25 years of the Black/White achievement gap. *British Educational Research Journal*, *43*(5), 848–874.

Gillborn, D., Rollock, N., Vincent, C., & Ball, S. J. (2012). "You got a pass, so what more do you want?": race, class and gender intersections in the educational experiences of the Black middle class. *Race Ethnicity and Education*, *15*(1), 121–139.

Girard, A. and Bastide, H. (1963). 'La stratification sociale et la démocratisation de l'enseignement'. *Population, 18*(3), 435-472.

Goldthorpe, J. H. (1996). Class analysis and the reorientation of class theory: The case of persisting differentials in educational attainment. *The British Journal of Sociology*, *47*(3), 481-505.

Goldthorpe, J. H., & Mills, C. (2008). Trends in Intergenerational Class Mobility in Modern Britain: Evidence From National Surveys, 1972—2005. *National Institute Economic Review, 205*(1), 83-100.

Goodman, A., Gregg, P., & Washbrook, E. (2011). Children's educational attainment and the aspirations, attitudes and behaviours of parents and children through childhood in the UK. *Longitudinal and Life Course Studies*, *2*(1), 1–18.

Gorard, S. & See, B. H. (2009). The impact of SES on participation and attainment in science. *Studies in Science Education, 45*(1), 93–129.

Gorard, S., See, B. H., & Davies, P. (2012). *The impact of attitudes and aspirations on educational attainment and participation.* York: Joseph Rowntree Foundation.

Gottburgsen, A., Gross, C. (2012). Welchen Beitrag leistet „Intersektionalität" zur Klärung von Kompetenzunterschieden bei Jugendlichen? *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 52,* 86–110.

Gottfredson, L. S. (2004). Intelligence: is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, *86*(1), 174–199.

Goyette, K. A., & Mullen, A. L. (2006). Who studies the arts and sciences? Social background and the choice and consequences of undergraduate field of study. *The Journal of Higher Education*, *77*(3), 497–538.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206–213.

Greenwood, C., Harrison, M. & Vignoles, A. (2011). *The labour market value of STEM qualifications and occupations. An analysis for the Royal Academy of Engineering.* Retrieved from: https://www.raeng.org.uk/publications/reports/the_labour_market_value_of_stem

Gregg, P., Jonsson, J. O., Macmillan, L., & Mood, C. (2017). The role of education for intergenerational income mobility: A comparison of the United States, Great

Britain, and Sweden. *Social Forces*, *96*(1), 121–152.

Gross, C, Meyer, H-D & Hadjar, A. (2016a). Theorising the impact of education systems on inequalities in: Hadjar, A & Gross, C. (Eds) *Education Systems and Inequalities*. Bristol: Policy Press.

Gross, C, Gottburgsen, A. & Phoenix, A. (2016b) Education systems and intersectionality in: Hadjar, A & Gross, C. (Eds) *Education Systems and Inequalities*. Bristol: Policy Press.

Gugushvili, A., Bukodi, E., & Goldthorpe, J. H. (2017). The direct effect of social origins on social mobility chances: "glass floors" and "glass ceilings" in britain. *European Sociological Review*, *33*(2), 305–316.

Gutman, L., & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people*. London: Education Endowment Foundation.

Hankivsky, O. (2011). *Health inequities in Canada: Intersectional frameworks and practices*. Vancouver/Toronto: UBC Press.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, *46*(3), 607–668.

Harnois, C. E. (2013). *Feminist measures in survey research*. Thousand Oaks: Sage.

Heath, A. F., Rothon, C., & Kilpi, E. (2008). The second generation in Western Europe: Education, unemployment, and occupational attainment. *Annual Review of Sociology*, *34*(1), 211–235.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900-1902.

Henderson, M., Sullivan, A., Anders, J. and Moulton, V. (2016). *Social Class, Gender and Ethnic Differences in Subjects Taken at Age 14.* CLS working paper 2016/6. London: Centre for Longitudinal Studies.

Hill Collins, P. (2002). *Black feminist thought: Knowledge, consciousness and the politics of empowerment*. London: Routledge.

HM Government. (2017). *Building our Industrial Strategy. Green Paper.* Retrieved from https://beisgovuk.citizenspace.com/strategy/industrial-strategy/supporting_documents/buildingourindustrialstrategygreenpaper.pdf

Hobbs, G. & Vignoles, A. (2007). *Is free school meal status a valid proxy for socio-economic status (in schools research)?* London: London School of Economics and Political Science.

Howard, K. A., Carlstrom, A. H., Katz, A. D., Chew, A. Y., Ray, G. C., Laine, L., & Caulum, D. (2011). Career aspirations of youth: Untangling race/ethnicity, SES, and gender. *Journal of Vocational Behavior*, *79*(1), 98-109.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60(6)*, 581.

Iannelli, C., & Duta, A. (2018). Inequalities in school leavers labour market outcomes: do school subject choices matter? *Oxford Review of Education, 44*(1), 56-74.

Ishida, H., Spilerman, S., & Su, K.-H. (1997). Educational credentials and promotion chances in japanese and american organizations. *Source American Sociological Review*, *62*(6), 866–882.

Jaccard, J. (2001). *Interaction effects in logistic regression*. Thousand Oaks, CA: Sage.

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*, *50*(3), 211–229.

Jacobs, J. A. (1996). Gender inequality and higher education. *Annual review of sociology*, *22*(1), 153-185.

Jacob, M., Klein, M., & Iannelli, C. (2015). The impact of social origin on graduates' early occupational destinations - an anglo-german comparison. *European Sociological Review*, *31*(4), 460–476.

Jäntti, M., Bratsberg, B., Røed, K., Raaum, O., Österbacka, E., Björklund, A., & Eriksson, T. (2006). American exceptionalism in a new light : a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States. *IZA Discussion Paper*, (1938), 1–40.

Jerrim, J., Chmielewski, A. K., & Parker, P. (2015). Socioeconomic inequality in access to high-status colleges: A cross-country comparison. *Research in Social Stratification and Mobility*, *42*, 20–32.

Jerrim, J., & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *176*(4), 887–906.

Jerrim, J., & Vignoles, A. (2015). University access for disadvantaged children: a comparison across countries. *Higher Education*, *70*(6), 903–921.

Johnson, W., McGue, M., & Iacono, W. G. (2006). Genetic and environmental influences on academic achievement trajectories during adolescence.

Developmental Psychology, *42*(3), 514–532.

Joint Council for Qualifications. (2014). *GCE results 2014*. Retrieved from https://www.jcq.org.uk/examination-results/a-levels/2014/a-as-and-aea-results-summer-2014

Jones, P., & Elias, P. (2005). *Science, engineering and technology and the UK's ethnic minority population*. London: Royal Society.

Kane, T. J. (1994). College entry by blacks since 1970: the role of college costs, family background, and the returns to education. *Journal of Political Economy*, *102*(5), 878-911.

Kenway, P. & Palmer, G. (2007). *Poverty among ethnic groups: How and why does it differ?* York: Joseph Rowntree Foundation.

Khattab, N. (2015). Students' aspirations, expectations and school achievement: what really matters? *British Educational Research Journal*, *41*(5), 731–748.

Kingdon, G. & Cassen, R. (2010). Ethnicity and low achievement in English schools, *British Educational Research Journal, 36*(3), 403–431.

Kirkup, G., Zalevski, A., Maruyama, T. & Batool, I. (2010). *Women and men in science, engineering and technology: The UK statistics guide 2010.* Bradford: UK Resource Centre for Women in Science, Engineering and Technology.

Kohn, M. L., & Schooler, C. (1983). *Work and personality: An inquiry into the impact of social stratification.* Norwood, NJ: Ablex.

Kolenikov, S. & Angeles, G. (2004). *The use of discrete data in PCA: Theory, simulations, and applications to socioeconomic indices*. Chapel Hill, NC: Carolina Population Centre, University of North Carolina.

Kulik, L. (2002). The impact of social background on gender-role ideology: Parents' versus children's attitudes, *Journal of Family Issues, 23*(1), 53–73.

Kutnick, P., Sebba, J., Blatchford, P., Galton, M., Thorp, J., MacIntyre, H., & Berdondini, L. (2005). The effects of pupil grouping: Literature review. Retrieved from: http://dera.ioe.ac.uk/18143/1/RR688.pdf

Lareau, A. (2000). *Home advantage: Social class and parental intervention in elementary education.* Lanham, MD: Rowman & Littlefield Publishers.

Lareau, A. (2003). *Unequal childhoods: Race, class and family life*. Berkeley: University of California Press.

Lareau, A. (2006). Unequal childhoods: class, race and family life. *American Journal of Sociology*, *112*, 635–636.

Laurison, D., & Friedman, S. (2016). The class pay gap in higher professional and managerial occupations. *American Sociological Review*, *81*(4), 668–695.

Leppel, K., Williams, M. L., & Waldauer, C. (2001). The impact of parental occupation and socioeconomic status on choice of college major. *Journal of Family and Economic Issues*, *22*(4), 373–394.

Lessard-Phillips, L., & Li, Y. (2017) Social stratification of education by ethnic minority groups over generations in the UK. *Social Inclusion*, *5*(1), 2183–2803.

Linn, M. C. & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher, 18*(8), 17–27.

Little, R. J., & Rubin, D. (1987). *Statistical analysis with missing data. Series in probability and mathematical statistics*. New York: John Wiley & Sons

Loehlin, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist*, *44*(10), 1285–1292.

Lucas, S. R. (2001). Effectively maintained inequality: education transitions, track mobility, and social background effects. *American Journal of Sociology*, *106*(6), 1642–1690.

Lutz, H., Herrera Vivar, M. T., & Supik, L. (2016). *Framing intersectionality debates on a multi-faceted concept in gender studies*. Farnham: Ashgate.

Lyons, T. (2006). Different countries, same science classes: students' experiences of school science in their own words. *International Journal of Science Education, 28*(6), 591-613.

Lyons, T., & Quinn, F. (2010). *Choosing science: Understanding the declines in senior high school science enrolments.* Retrieved from https://eprints.qut.edu.au/68725/1/Choosing_Science.pdf

Ma, Y. (2009). Family socioeconomic status, parental involvement, and college major choices— gender, race/ethnic, and nativity patterns. *Sociological Perspectives*, *52*(2), 211–234.

Mann, A., & DiPrete, T. (2016). The consequences of the national math and science performance environment for gender differences in STEM aspirations. *Sociological Science*, *3*, 568–603.

Manski, C. F. (1993). Adolescent econometricians: How do youth infer the returns to schooling? Studies of supply and demand in higher education. In: Clotfelter, C. T., & Rothschild, M. (Ed.) *Studies of supply and demand in higher education* (pp. 43-60). Chicago: University of Chicago Press.

Marien, S., Hooghe, M., & Quintelier, E. (2010). Inequalities in non-institutionalised forms of political participation: A multi-level analysis of 25 countries. *Political Studies*, *58*(1), 187–213.

Marks, G. N. (2005a). Accounting for immigrant non-immigrant differences in reading and mathematics in twenty countries. *Ethnic and Racial Studies*, *28*(5), 925–946.

Marks, G. N. (2005b) Cross-national differences and accounting for social class inequalities in education. *International Sociology*, *20*(4), 483–505.

Mau, W. C., Domnick, M. & Ellsworth, R. A. (1995). Characteristics of female students who aspire to science and engineering or homemaking occupations. *Career Development Quarterly, 43*(4), 323.

McCall, L. (2001). *Complex inequality: gender, class, and race in the new economy*. New York: Routledge.

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society,* 30(3), 1771-1800.

Melhuish, E. (2014). The impact of early childhood education and care on improved wellbeing. In *"If you could do one thing..." Nine local actions to reduce health inequalities* (pp 33-43). London: British Academy.

Mensah, F. K., & Kiernan, K. E. (2010). Gender differences in educational attainment: influences of the family environment. *British Educational Research Journal*, *36*(2), 239-260.

Mirza, H. S. & Joseph, C. (2013). *Black and postcolonial feminisms in new times: Researching educational inequalities*. London: Routledge.

Moakler, M, W., & Kim, M. M. (2014). College major choice in stem : revisiting confidence and demographic factors. *The Career Development Quarterly*, *62(2)*, 128–142.

Modood, T. (2003). Ethnic differentials in educational performance. In: Mason, D. (Ed.) *Explaining Ethnic Differences* (pp 53-68). Bristol: Policy Press.

Montt, G. (2011). Cross-national differences in educational achievement inequality. *Sociology of Education*, *84*(1), 49–68.

Mortimer, J. T., Lorence, J., & Kumka, D. S. (1986). *Work, family, and personality: Transition to adulthood*. Westport, CT: Ablex Publishing.

Mostafa, T., & Wiggins, R. D. (2014). Handling attrition and non-response in the 1970 British Cohort Study. *CLS Working Paper 2014/2*.

Mostafa, T., & Wiggins, R. (2015). The impact of attrition and non-response in birth cohort studies: a need to incorporate missingness strategies. *Longitudinal and Life Course Studies, 6*(2), 131-146.

Mujtaba, T. & Reiss, M. J. (2012). *Factors affecting whether students in England choose to study physics once the subject is optional*. London: Institute of

Education.

National Audit Office. (2010). *Educating the next generation of scientists.* London: The Stationary Office.

Naylor, R., Smith, J., & Telhaj, S. (2016). Graduate returns, degree class premia and higher education expansion in the UK. *Oxford Economic Papers*, *68*(2), 525–545.

Noyes, A. (2009). Exploring social patterns of participation in university-entrance level mathematics in England, Research in Mathematics Education, 11(2), 167–183.

OECD (2012). *Education at a glance 2012.* Paris: OECD Publications.

OECD. (2015). *The ABC of gender equality in education: Aptitude, Behaviour, Confidence*. Paris: OECD.

Pallas A.M. (2000). The effects of schooling on individual lives. In: Hallinan, M. T. (Ed.) *Handbooks of sociology and social research* (pp 499-525). Boston, MA: Springer.

Pampaka, M., Williams, J. & Hutcheson, G. (2012a). Measuring students' transition into university and its association with learning outcomes. *British Educational Research Journal, 38*(6), 1041–1071.

Pampaka, M., Williams, J., Hutcheson, G., Wake, G., Black, L., Davis, P. & Hernandez-Martinez, P. (2012b). The association between mathematics pedagogy and learners' dispositions for university study. *British Educational Research Journal, 38*(3), 473–496.

Paulsen, M. B. (2001). The economics of human capital and investment in higher education. In: Paulsen, M. B., & Smart, J. C. (Ed.) *The Finance of Higher Education: Theory, Research, Policy, and Practice* (pp 55–94). New York: Algora Publishing.

Peters, M. A. (2015). Why is my curriculum white? *Educational Philosophy and Theory, 47*(7), 641–646.

Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European Sociological Review*, *24*(5), 543-565.

Rimfeld, K., Ayorech, Z., Dale, P. S., Kovas, Y., & Plomin, R. (2016). Genetics affects choice of academic subjects as well as achievement. *Scientific Reports*, *6*, Article number: 26373. (Online)

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, *4*, 227–241.

Runnymede Trust. (2014). *When education isn't enough: Labour market outcomes of* *ethnic minority graduates at elite universities.* Retrieved from: https://www.runnymedetrust.org/uploads/WhenEducationIsntEnough.pdf

Rutter, M. (2006). *Genes and behavior: nature-nurture interplay explained.* Oxford, UK: Blackwell Publishers

Sa, F. (2017). The effect of university fees on applications, attendance and course choice: evidence from a natural experiment in the uk. *Economica*. (Online First).

Sacker, A., Schoon, I., & Bartley, M. (2002). Social inequality in educational achievement and psychosocial adjustment throughout childhood: magnitude and mechanisms. *Social science & medicine, 55*(5), 863-880.

Schoon, I., & Lyons-Amos, M. (2017). A socio-ecological model of agency: The role of structure and agency in shaping education and employment transitions in England. *Journal of Longitudinal and Lifecourse Studies, 8*(1), 35-56.

Schulz, W. H. (2005). *Mathematics self-efficacy and student expectations- results from PISA 2003. Annual meeting of the American Educational Research Association: Demography and Democracy in the Era of Accountability, (Montreal, Apr 11-15, 2005)*. Retrieved from: https://eric.ed.gov/?id=ED490044

Scott, J. (2010). Quantitative methods and gender inequalities. *International Journal of Social Research Methodology*, *13*(3), 223–236.

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in stem fields. *Sex Roles*, *66*(3–4), 175–183.

Sharples, J., Slavin, R., Chambers, B., & Sharp, C. (2011). *Effective classroom strategies for closing the gap in educational achievement for children and young people living in poverty, including white working-class boys*. London: C4EO.

Shavit, Y., & Blossfeld, H. P. (1993). *Persistent inequality: changing educational attainment in thirteen countries*. Boulder, CO: Westview Press.

Shavit, Y., & Muller, W. (1998). *From school to work. A comparative study of educational qualifications and occupational destinations*. Cary, NC: Oxford University Press.

Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2014). Students' intentions to study non-compulsory mathematics: the importance of how good you think you are. *British Educational Research Journal, 41*(3), 462-488.

Shipman, T., & Griffiths, S. (2018, February 18). Education secretary Damian Hinds:

students to get cheaper places at university. *The Sunday Times*. Retrieved from https://www.thetimes.co.uk/article/education-secretary-damian-hinds-students-to-get-cheaper-places-at-university-htqxq326z

Smith, E. (2011). Women into science and engineering? Gendered participation in higher education STEM subjects. *British Educational Research Journal*, *37*(6), 993–1014.

Smith, E., & White, P. (2017). A "great way to get on"? The early career destinations of science, technology, engineering and mathematics graduates. *Research Papers in Education*, *32*(2), 231–253.

Spierings, N. (2012). The inclusion of quantitative techniques and diversity in the mainstream of feminist research. *European Journal of Women's Studies*, *19*(3), 331-347.

Spilerman, S., & Lunde, T. (1991). Features of educational attainment and job promotion prospects. *American Journal of Sociology*, *97*(3), 689.

Stahl, G. (2017). Aspiration paradoxes: working-class student conceptions of power in "engines of social mobility." *International Journal of Qualitative Studies in Education*. (Online First).

Strand, S. (2007). Minority ethnic pupils in the Longitudinal Study of Young People in England (LSYPE). London: Department for Children, Schools and Families.

Strand, S. (2011). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal, 37*(2), 197–229.

Strand, S. (2014a). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'getting it' for the white working class. *Research Papers in Education, 29*(2), 131–171.

Strand, S. (2014b). School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11. *Oxford Review of Education*, *40*(2), 223–245.

Sui-Chu, E. H., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of education, 69*(2), 126-141.

Sullivan, A., Parsons, S., Green, F., Wiggins, R. D., & Ploubidis, G. (2017). The path from social origins to top jobs: social reproduction via education. *The British Journal of Sociology, 69(3)*, 776-798.

Sullivan, A., Parsons, S., Green, F., Wiggins, R. D., Ploubidis, G. (2018). Elite universities, fields of study and top salaries: Which degree will make you rich? *British educational research journal, 44(4), 663-680.*

The Royal Society. (2008). *Science and mathematics education, 14–19. A "state of the* *nation" report on the participation and attainment of 14–19 year olds in science and mathematics in the UK, 1996–2007*. London: The Royal Society.

TISME. (2013). *What influences participation in science and mathematics? A briefing paper from the Targeted Initiative on Science and Mathematics Education (TISME).* Retrieved from: https://www.kcl.ac.uk/sspp/departments/education/research/aspires/TISME-briefing-paper-March-2013.pdf

Trahan, A. (2011). Qualitative research and intersectionality. *Critical Criminology*, 19 (1), 1–14.

Triventi, M. (2013). The role of higher education stratification in the reproduction of social inequality in the labor market. *Research in Social Stratification and Mobility*, *32*(1), 45–63.

UCAS. (2016). *2016 UCAS End of Cycle Report.* Retrieved from https://www.ucas.com/file/86541

Trusty, J., Robinson, C. R., Plata, M., & Ng, K. M. (2000). Effects of gender, socioeconomic status, and early academic performance on postsecondary educational choice. *Journal of Counseling and Development, 78*(4), 463–472.

van de Werfhorst, H. G. (2017). Gender segregation across fields of study in post-secondary education: trends and social differentials. *European Sociological Review*, *33*(3), 449–464.

van de Werfhorst, H. G., & Hofstede, S. (2007). Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared. *British Journal of Sociology, 58*(3), 391-415.

van de Werfhorst, H. G., & Luijkx, R. (2010). Educational field of study and social mobility: disaggregating social origin and education. *Sociology*, *44*(4), 695–715.

van de Werfhorst, H. G., & Mijs, J. J. B. (2010) Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, *36*, 407–28.

van de Werfhorst, H. G., Sullivan, A., & Cheung, S. Y. (2003). Social class, ability and choice of subject in secondary and tertiary education in britain. *British Educational Research Journal*, *29*(1), 41–62.

van Langen, A., & Dekkers, H. (2005). Cross-national differences in participating in tertiary science, technology, engineering and mathematics education. *Comparative Education*, *41*(3), 329–350.

Walker, I., & Zhu, Y. (2001). *The returns to education: evidence from the labour force* *surveys.* Retrieved from: http://dera.ioe.ac.uk/4656/1/RR313.pdf

Walker, I., & Zhu, Y. (2008). The college wage premium and the expansion of higher education in the UK. *Scandinavian Journal of Economics*, *110*(4), 695–709.

Walker, I., & Zhu, Y. (2011). Differences by degree: Evidence of the net financial rates of return to undergraduate study for England and Wales. *Economics of Education Review*, *30*(6), 1177–1186.

Walker, I., & Zhu, Y. (2013). *The impact of university degrees on the lifecycle of earnings: some further analysis. Department for Business Innovation and Skills.* Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/229498/bis-13-899-the-impact-of-university-degrees-on-the-lifecycle-of-earnings-further-analysis.pdf

Walker, S. P., Wachs, T. D., Grantham-Mcgregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., Richter, L. et al. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, *378*(9799), 1325–1338.

Wang, M. T., Eccles, J. S. & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24*(5), 770–775.

Warikoo, N. & Carter, P. (2009). Cultural explanations for racial and ethnic stratification in academic achievement: A call for a new and improved theory. *Review of Educational Research, 79*(1), 366–394.

Weber, L. (2001). *Understanding race, class, gender, and sexuality: a conceptual framework*. Boston: McGraw-Hill.

Webber, D. A. (2014). The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors. *Labour Economics*, *28*, 14–23.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-vaue theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68-81.

Winterbotham, M. (2014). *UK Commission's Employer Skills Survey 2013: UK results* London: UK Commission.

Wynarczyk, P. & Hale, S. (2008). *Improving take up of science and technology subjects in schools and colleges: A synthesis review. Report prepared for the*

Economic and Social Research Council (ESRC) 'Science in Society' Team and the Department for Children, Schools and Families (DCSF). Newcastle: Small Enterprise Research Unit, Newcastle University Business School.

Appendix A: Variables used in multiple imputation models

Table A1: Variables used in multiple imputations for Next Steps data

Year	Variables
2004	Parents' income
	Whether independently educated
	Parents' qualifications (mothers and fathers)
	Parents self-reported financial stability
	Housing tenure
	Parents' social class
	Motivation at school
	Enjoyment of English, Science and Maths
	Self concept in English, Science and Maths
	Ethnicity
	Sex
2005	Locus of control
	Whether engage in risky behaviour
2010	Whether studying in higher education or university
	Whether studying in a Russell group university
	Subject studied at university

Table A2: Variables used in multiple imputations for NLSY79 data

Year	Variables
1979	Ethnicity
	Country of birth (migrant)
	Age
	Urban/ Rural
	Number of siblings
	Whether privately educated
	Housing tenure
	Age of mother at birth
	Birthplace of mother (us or migrant)
	Did any household member have a library card at age 14
	Does health limit moderate activities
	Parent's occupational class (Duncan socio-economic index)
	Family income
	Parent's education (years within levels)
	No father present (biological or step)
	Gender views
	AFQT (cognitive ability)
Later	Biannual income from 1988-2012
	Marital status in 2004

Table A3: Variables used in multiple imputations for BCS70 data

Age	Variables
0	Birthweight
	Birth order
	Parity
	Age of mother
	Marital status
	Parent social class
	Parent's age left full time education
	Frequency parent read to child
	Cognition
5	Ever breastfed
	Overcrowded home
	Home ownership
	Ethnicity
	Parent's qualifications
	Whether father (bio or step) is not present in the household
	Rutter behaviour score
	Cognition
10	Ever received state benefits
	Free school meals
	Family income
	Privately educated
	Whether has a disability that interferes with daily life
	Urban/ rural
	Parents interest in child's education
	Whether father (bio or step) is not present in the household
	Rutter behaviour score
	Cognition
16	Whether father (bio or step) is not present in the household
	Rutter behaviour score
	Cognition
26	Yearly income
	Gender views
30	Yearly income
	Gender views
34	Yearly income
38	Yearly income
42	Yearly income
	Social class
	Marital status
	Average UCAS points of university attended

Appendix B: Additional material for Chapter 5

Intersections between students' characteristics and subject choices in chapter five were represented using interaction terms, however sub-group analysis was also conducted to confirm that relationships remained, and indeed ran in the direction suggested by the interaction terms. This follows Ai and Norton (2003), who pointed out that interaction terms are not always easy to interpret within models. The analysis using Next Steps data was published as a working paper (Codiroli, 2015). Note that for multinomial regressions of subject choice, arts and humanities subjects are the base category, not STEM subjects (as in chapter five).

Appendix B1: Results using Next Steps data

Table B1.1: Results of logistic regression of choice of at least one STEM A-level, marginal effects are shown with standard errors in parenthesis

	Female		Male	
	Model 1	Model 2	Model 1	Model 2
Ethnicity				
White				
Mixed	0.565* (0.323)	0.613* (0.362)	0.009 (0.290)	0.102 (0.495)
Indian	1.080*** (0.171)	1.387*** (0.204)	0.861*** (0.198)	1.429*** (0.281)
Pakistani	1.097*** (0.247)	1.655*** (0.255)	0.848*** (0.247)	1.677*** (0.342)
Bangladeshi	0.088 (0.245)	0.309 (0.296)	0.579* (0.310)	0.974*** (0.364)
Black Caribbean	0.941*** (0.363)	1.563*** (0.411)	-0.231 (0.424)	1.059** (0.440)
Black African	-0.028 (0.328)	0.570 (0.381)	-0.001 (0.332)	0.810 (0.506)
Other	1.294*** (0.328)	1.175*** (0.339)	1.241*** (0.439)	1.548*** (0.567)
Social class				
Higher managerial	0.550* (0.284)	0.334 (0.325)	0.192 (0.292)	-0.388 (0.350)
Lower managerial	0.400 (0.272)	0.317 (0.309)	-0.281 (0.276)	-0.842** (0.334)
Intermediate	0.431 (0.322)	0.318 (0.374)	-0.101 (0.322)	-0.204 (0.386)
Small employer	0.207 (0.298)	0.423 (0.341)	-0.200 (0.296)	-0.452 (0.360)
Lower supervisor	-0.079 (0.329)	0.104 (0.381)	-0.209 (0.330)	-0.463 (0.401)
Semi-routine	-0.131 (0.334)	0.069 (0.378)	-0.050 (0.317)	-0.286 (0.392)
Reference: Routine				
Unemployed	0.050 (0.386)	0.297 (0.454)	-0.736 (0.485)	-0.858 (0.544)
Mothers has a degree of higher	0.183 (0.147)	-0.167 (0.173)	0.263* (0.149)	-0.142 (0.176)
Fathers has a degree or higher	0.654*** (0.149)	0.404** (0.176)	0.311** (0.155)	0.298* (0.178)
Income	-0.002 (0.005)	-0.004 (0.005)	0.010** (0.005)	0.006 (0.005)
Independent School	0.453** (0.186)	-0.211 (0.372)	-0.144 (0.197)	-0.136 (0.397)
Prior attainment				
GCSE score		1.284*** (0.103)		1.371*** (0.108)
KS2 Math		0.789*** (0.104)		0.821*** (0.101)
KS2 Science		0.076 (0.103)		0.311*** (0.105)
KS2 English		-0.529*** (0.097)		-0.709*** (0.099)
N (STEM A-level)	2275 (722)		1853 (872)	

*p < 0.10, ** p < 0.05, *** p < 0.01

Table B1.2: Results of logistic regression of choice of at least one STEM A-level, stratified by students SEP, marginal effects are shown with standard errors in parenthesis

	Low SEP	2nd SEP	High SEP
Male	0.124***	0.126***	0.154***
	(0.029)	(0.031)	(0.037)
BME	0.183***	0.231***	0.271***
	(0.030)	(0.040)	(0.064)
GCSE	0.244***	0.247***	0.327***
	(0.023)	(0.024)	(0.031)
KS2 Math	0.128***	0.197***	0.173***
	(0.021)	(0.024)	(0.031)
KS2 Science	0.041*	0.010	0.068**
	(0.021)	(0.026)	(0.029)
KS2 English	-0.098***	-0.117***	-0.185***
	(0.020)	(0.023)	(0.029)
N (N STEM A-level)	1,463 (482)	1,328 (492)	1,337 (620)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B1.3: Results of multinomial logistic regression of degree choice for female students, marginal effects are shown with standard errors in parenthesis

	STEM						SLB					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
Ethnicity												
White												
Mixed	0.004	(0.315)	0.042	(0.312)	-0.102	(0.354)	0.621*	(0.340)	0.659*	(0.338)	0.665*	(0.343)
Indian	0.808***	(0.234)	0.811***	(0.242)	0.300	(0.250)	1.322***	(0.226)	1.287***	(0.239)	1.250***	(0.243)
Pakistani	1.472***	(0.316)	1.526***	(0.317)	0.993***	(0.318)	1.800***	(0.294)	1.859***	(0.312)	1.788***	(0.314)
Bangladeshi	0.701**	(0.343)	0.770**	(0.337)	0.570	(0.348)	1.370***	(0.362)	1.403***	(0.362)	1.372***	(0.360)
Black Caribbean	0.167	(0.444)	0.382	(0.543)	-0.102	(0.616)	0.240	(0.324)	0.317	(0.348)	0.291	(0.353)
Black African	0.927***	(0.340)	1.049***	(0.346)	0.932**	(0.392)	1.767***	(0.318)	1.778***	(0.340)	1.764***	(0.339)
Other	1.040***	(0.367)	0.981***	(0.369)	0.563	(0.380)	0.851**	(0.399)	0.876**	(0.411)	0.863**	(0.412)
Social class												
Higher managerial	0.079	(0.300)	0.067	(0.303)	-0.160	(0.318)	-0.347	(0.304)	-0.288	(0.305)	-0.290	(0.305)
Lower managerial	0.006	(0.282)	0.010	(0.284)	-0.128	(0.302)	-0.414	(0.279)	-0.386	(0.279)	-0.373	(0.280)
Intermediate	-0.449	(0.346)	-0.496	(0.352)	-0.807**	(0.384)	-0.413	(0.340)	-0.412	(0.339)	-0.406	(0.340)
Small employer	-0.190	(0.316)	-0.141	(0.317)	-0.320	(0.338)	-0.419	(0.312)	-0.354	(0.312)	-0.344	(0.313)
Lower supervisor	0.010	(0.333)	0.037	(0.337)	0.050	(0.354)	-0.031	(0.324)	-0.036	(0.321)	-0.027	(0.322)
Semi-routine	0.107	(0.335)	0.195	(0.332)	0.174	(0.361)	-0.694**	(0.342)	-0.670*	(0.346)	-0.643*	(0.348)
Reference: Routine												
Unemployed	-0.446	(0.428)	-0.343	(0.430)	-0.376	(0.447)	-1.158***	(0.439)	-1.121**	(0.442)	-1.053**	(0.436)
Mothers has a degree	-0.054	(0.159)	-0.118	(0.166)	-0.117	(0.177)	-0.513**	(0.204)	-0.458**	(0.209)	-0.454**	(0.208)
Fathers has a degree	0.315*	(0.163)	0.230	(0.166)	0.056	(0.177)	-0.130	(0.202)	-0.082	(0.207)	-0.118	(0.205)
Income	-0.004	(0.005)	-0.005	(0.005)	-0.005	(0.005)	-0.001	(0.005)	0.000	(0.005)	0.000	(0.005)
Independent School	-0.705*	(0.367)	-0.726*	(0.374)	-0.630	(0.430)	0.120	(0.364)	0.170	(0.371)	0.161	(0.375)
Prior attainment												
GCSE score			0.140	(0.102)	-0.205**	(0.095)			-0.172*	(0.088)	-0.184*	(0.098)
KS2 Math			0.467***	(0.099)	0.221**	(0.109)			0.458***	(0.102)	0.445***	(0.102)
KS2 Science			-0.008	(0.093)	-0.072	(0.100)			-0.137	(0.100)	-0.135	(0.099)
KS2 English			-0.292***	(0.096)	-0.134	(0.101)			-0.132	(0.098)	-0.122	(0.099)
One + STEM A-levels					2.292***	(0.164)					0.152	(0.200)
Two + STEM A-levels					-0.002	(0.152)					0.025	(0.152)
N						2,289 (781/ 659)						

* p < 0.10, ** p < 0.05, *** p < 0.01

Table B1.4: Results of multinomial logistic regression of degree choice for male students, marginal effects are shown with standard errors in parenthesis

	STEM						SLB					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
Ethnicity												
White												
Mixed	-0.092	(0.353)	-0.117	(0.366)	0.019	(0.356)	-0.324	(0.438)	-0.326	(0.434)	-0.313	(0.431)
Indian	0.545*	(0.281)	0.785**	(0.320)	0.376	(0.330)	0.839***	(0.296)	0.913***	(0.335)	0.845**	(0.342)
Pakistani	1.317***	(0.432)	1.750***	(0.411)	1.289***	(0.436)	1.915***	(0.431)	2.104***	(0.412)	2.028***	(0.429)
Bangladeshi	1.189***	(0.408)	1.434***	(0.435)	1.213***	(0.428)	1.623***	(0.418)	1.724***	(0.427)	1.700***	(0.427)
Black Caribbean	-0.056	(0.416)	0.628	(0.445)	0.416	(0.454)	0.519	(0.427)	0.877**	(0.435)	0.825*	(0.439)
Black African	-0.202	(0.332)	0.109	(0.364)	-0.019	(0.384)	0.371	(0.365)	0.478	(0.387)	0.460	(0.395)
Other	0.417	(0.450)	0.439	(0.453)	0.090	(0.408)	-0.017	(0.528)	-0.026	(0.554)	-0.128	(0.552)
Social class												
Higher managerial	-0.498	(0.361)	-0.851**	(0.360)	-0.812**	(0.377)	0.155	(0.419)	-0.005	(0.411)	0.082	(0.417)
Lower managerial	-0.794**	(0.343)	-1.096***	(0.338)	-0.866**	(0.353)	-0.022	(0.392)	-0.183	(0.380)	-0.060	(0.391)
Intermediate	-1.108***	(0.391)	-1.228***	(0.391)	-1.346***	(0.404)	-0.526	(0.457)	-0.612	(0.446)	-0.573	(0.453)
Small employer	-0.688*	(0.371)	-0.896**	(0.366)	-0.877**	(0.383)	0.282	(0.419)	0.145	(0.405)	0.217	(0.413)
Lower supervisor	-1.060***	(0.406)	-1.277***	(0.406)	-1.283***	(0.442)	0.053	(0.441)	-0.064	(0.429)	0.004	(0.438)
Semi-routine	-0.984**	(0.393)	-1.073***	(0.389)	-1.107***	(0.418)	-0.143	(0.436)	-0.149	(0.418)	-0.099	(0.430)
Reference: Routine												
Unemployed	-0.592	(0.586)	-0.860	(0.600)	-0.761	(0.535)	0.084	(0.622)	-0.060	(0.610)	0.009	(0.605)
Mothers has a degree	-0.149	(0.168)	-0.338*	(0.174)	-0.391**	(0.185)	-0.198	(0.203)	-0.310	(0.207)	-0.330	(0.209)
Fathers has a degree	-0.172	(0.178)	-0.272	(0.183)	-0.380*	(0.196)	-0.304	(0.211)	-0.381*	(0.218)	-0.410*	(0.219)
Income	0.009*	(0.005)	0.006	(0.005)	0.003	(0.005)	0.001	(0.006)	-0.002	(0.006)	-0.002	(0.006)
Independent School	-0.090	(0.361)	-0.039	(0.378)	0.146	(0.337)	0.359	(0.402)	0.366	(0.428)	0.385	(0.426)
Prior attainment												
GCSE score			0.448***	(0.099)	0.103	(0.117)			0.167	(0.107)	0.027	(0.109)
KS2 Math			0.274***	(0.101)	0.170	(0.117)			0.208*	(0.117)	0.011	(0.108)
KS2 Science			0.264***	(0.101)	-0.106	(0.114)			-0.085	(0.111)	0.175	(0.110)
KS2 English			-0.349***	(0.096)	0.028	(0.108)			-0.015	(0.110)	-0.147	(0.101)
One + STEM A-levels					2.099***	(0.171)					0.452**	(0.200)
Two + STEM A-levels					-0.046	(0.179)					0.114	(0.707)
N	1,846 (806/ 496)											

* p < 0.10, ** p < 0.05, *** p < 0.01

Table B1.5: Results of multinomial logistic regression of choice of studying STEM over arts and humanities by students SEP. Marginal effects are shown with standard errors in parenthesis.

	Low SEP	2nd SEP	High SEP
Male	0.072**	0.063**	0.027
	(0.032)	(0.032)	(0.032)
BME	0.031	0.064	0.100**
	(0.031)	(0.040)	(0.046)
GCSE	0.084***	0.047	0.070***
	(0.020)	(0.031)	(0.025)
KS2 Math	0.068***	0.068***	0.028
	(0.023)	(0.024)	(0.028)
KS2 Science	-0.007	0.031	0.104***
	(0.021)	(0.024)	(0.027)
KS2 English	-0.057***	-0.052**	-0.110***
	(0.021)	(0.024)	(0.026)
N (N STEM degree)	1,539 (562)	1,289 (485)	1,307 (540)

* p < 0.10, ** p < 0.05, *** p < 0.01

Table B1.6: Results of multinomial logistic regression of choice of studying SLB over arts and humanities by students SEP. Marginal effects are shown with standard errors in parenthesis.

	Low SEP	2nd SEP	High SEP
Male	-0.060**	0.017	0.033
	(0.030)	(0.030)	(0.027)
BME	0.128***	0.100***	0.068**
	(0.026)	(0.033)	(0.030)
GCSE	-0.039**	-0.028	-0.025
	(0.018)	(0.020)	(0.019)
KS2 Math	0.017	0.054**	0.023
	(0.021)	(0.023)	(0.021)
KS2 Science	0.015	-0.085***	-0.054***
	(0.020)	(0.022)	(0.021)
KS2 English	0.005	0.030	0.020
	(0.020)	(0.022)	(0.018)
N (N SLB degree)	1,539 (513)	1,289 (348)	1,307 (294)

* p < 0.10, ** p < 0.05, *** p < 0.01

Appendix B2: Results using HESA-NPD data

Table B2.1: Multinomial logistic regression results predicting subject choice based on student characteristics for students whose parents are highly educated

	Arts and Humanities		SLB	
Variables	RRR	SE	RRR	SE
Female	1.976***	(0.033)	1.320***	(0.024)
Ethnicity (Ref: White)				
Mixed	1.031	(0.050)	1.370***	(0.071)
Indian	0.243***	(0.014)	1.122***	(0.047)
Pakistani	0.258***	(0.023)	1.194***	(0.074)
Bangladeshi	0.466***	(0.079)	1.631***	(0.221)
Black Caribbean	0.940	(0.069)	1.807***	(0.129)
Black African	0.462***	(0.024)	1.813***	(0.076)
Other	0.456***	(0.023)	1.113**	(0.050)
Constant	0.331***	(0.009)	0.369***	(0.024)
Observations	79,032			

* p < 0.10, ** p < 0.05, *** p < 0.01

Table B2.2: Multinomial logistic regression results predicting subject choice based on student characteristics for students whose parents have lower education levels

	Arts and Humanities		SLB	
Variables	RRR	SE	RRR	SE
Female	1.925***	(0.034)	1.409***	(0.025)
Ethnicity (Ref: White)				
Mixed	0.918	(0.052)	1.306***	(0.075)
Indian	0.344***	(0.017)	1.424***	(0.050)
Pakistani	0.327***	(0.017)	1.447***	(0.054)
Bangladeshi	0.492***	(0.035)	1.835***	(0.099)
Black Caribbean	0.901	(0.068)	1.609***	(0.116)
Black African	0.514***	(0.036)	1.481***	(0.084)
Other	0.429***	(0.022)	1.112***	(0.048)
Constant	0.429***	(0.022)	0.390***	(0.012)
Observations	78,141			

* p < 0.10, ** p < 0.05, *** p < 0.01

Appendix C: Additional material for Chapter 7

Appendix C1: Field of study regressions with mothers and fathers education included separately[25]

Table C1.1 shows that in both countries fathers were more likely to have higher education levels than mothers, therefor variables combining both parents education would usually reflect the fathers education level. With that in mind, results for women in the US are very similar when including parent's education separately, with father's education explaining disparities rather than mother's education. For US men, however, parental education in now associated with field of study. Men whose mother's are less educated are more likely to study STEM, and those whose fathers are less educated are less likely to study STEM. For the UK sample parent's education is not significantly associated with field of study. The interpretability of these models is hindered by the high collinearity between parent's education levels.

Table C1.1: Mothers and fathers educational attainment (for graduates only) in BCS70 and NLSY79

Parent	Education level	UK %	US %
Mother	High education	30.30	39.83
	Low education	69.70	57.96
Father	High education	47.06	52.45
	Low education	51.56	43.07
	Father info is missing *and* he does not live in the household[26]	1.39	1.78

[25] High education and low education are defined in the same way as overall parent's education.
[26] Included as a dummy, not reported in tables due to very small samples

Table C1.2: Multinomial logistic regression predicting field of study for women (BCS70)

Base subject: OSSAH	STEM				LEM			
	1	2	3	4	1	2	3	4
Social background								
Mother has low education	0.911	0.883	0.914	0.994	0.987	0.989	0.979	0.946
	(0.177)	(0.172)	(0.180)	(0.199)	(0.210)	(0.214)	(0.214)	(0.208)
Father has low education	0.884	0.862	0.881	0.904	0.938	0.946	0.939	0.931
	(0.170)	(0.167)	(0.171)	(0.176)	(0.190)	(0.193)	(0.192)	(0.190)
FSM		2.025	2.183*	2.183*		2.806**	2.739**	2.752**
		(0.931)	(1.018)	(1.030)		(1.276)	(1.250)	(1.255)
Independent School		0.537*	0.525*	0.441**		1.258	1.266	1.357
		(0.197)	(0.193)	(0.164)		(0.409)	(0.411)	(0.451)
Father (bio or step) not present		0.748	0.752	0.786		0.665	0.666	0.651
		(0.232)	(0.235)	(0.244)		(0.233)	(0.233)	(0.230)
BME	1.332	1.202	1.351	1.368	1.773	1.568	1.514	1.508
	(0.726)	(0.661)	(0.752)	(0.778)	(0.963)	(0.871)	(0.855)	(0.844)
Cognitive ability (Mean of age 5 and 10 scores)			1.159	1.127			0.957	0.966
			(0.123)	(0.122)			(0.106)	(0.107)
Prestige of university								
Mean university acceptance scores				1.003***				0.998
				(0.001)				(0.001)
Observations	1054							

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C1.3: Multinomial logistic regression predicting field of study for women (NLSY79)

Base subject: OSSAH	STEM			LEM		
	1	2	3	1	2	3
Social background						
Mother has low education	0.840	0.849	0.902	1.195	1.231	1.317
	(0.181)	(0.185)	(0.199)	(0.275)	(0.286)	(0.310)
Father has low education	1.744***	1.793***	1.867***	1.741**	1.679**	1.772**
	(0.376)	(0.396)	(0.415)	(0.388)	(0.383)	(0.409)
Lowest decile income		0.920	0.938		0.562	0.574
		(0.471)	(0.483)		(0.313)	(0.321)
Attended a private school		1.656*	1.655		1.013	1.001
		(0.508)	(0.510)		(0.374)	(0.372)
Father (bio or step) not present		1.104	1.206		1.743*	1.940**
		(0.339)	(0.375)		(0.506)	(0.573)
Ethnicity						
Reference: Non-black, non-Hispanic						
Hispanic	1.865	1.854	2.301	1.521	1.462	1.908
	(0.997)	(0.996)	(1.265)	(0.872)	(0.841)	(1.121)
Black	1.744*	1.742	2.431**	1.623	1.617	2.374**
	(0.585)	(0.603)	(0.923)	(0.561)	(0.574)	(0.926)
Standardized AFQT			1.389**			1.479**
			(0.208)			(0.232)
Age	0.0850	0.125	0.177	1.831	1.517	2.239
	(0.147)	(0.221)	(0.314)	(3.476)	(2.932)	(4.348)
Age^2	1.028	1.024	1.020	0.992	0.994	0.990
	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)
Observations			659			

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C1.4: Multinomial logistic regression predicting field of study for men (BCS70)

Base subject: OSSAH	STEM				LEM			
	1	2	3	4	1	2	3	4
Social background								
Mother has low education	0.989	0.932	0.968	1.001	0.825	0.795	0.788	0.775
	(0.185)	(0.176)	(0.185)	(0.194)	(0.190)	(0.184)	(0.183)	(0.181)
Father has low education	0.949	0.907	0.970	0.993	1.047	0.997	0.979	0.966
	(0.164)	(0.161)	(0.175)	(0.180)	(0.236)	(0.228)	(0.228)	(0.226)
FSM		1.523	1.706	1.674		1.265	1.231	1.251
		(0.753)	(0.846)	(0.829)		(0.743)	(0.726)	(0.736)
Independent School		0.519**	0.498**	0.464**		0.663	0.667	0.691
		(0.165)	(0.159)	(0.150)		(0.238)	(0.239)	(0.251)
Father (bio or step) not present		0.695	0.686	0.707		1.055	1.061	1.044
		(0.225)	(0.224)	(0.233)		(0.371)	(0.373)	(0.367)
BME	2.471*	2.400*	3.166**	3.116**	1.106	1.097	1.027	1.034
	(1.212)	(1.188)	(1.622)	(1.602)	(0.762)	(0.763)	(0.742)	(0.746)
Cognitive ability (Mean of age 5 and 10 scores)			1.292**	1.271**			0.944	0.952
			(0.129)	(0.128)			(0.115)	(0.117)
Prestige of university								
Mean university acceptance scores				1.001				0.999
				(0.001)				(0.001)
Observations	1161							

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C1.5: Multinomial logistic regression predicting field of study for men (NLSY79)

Base subject: OSSAH	STEM			LEM		
	1	2	3	1	2	3
Social background						
Mother has low education	1.552*	1.498*	1.652**	0.938	0.958	0.952
	(0.367)	(0.361)	(0.405)	(0.242)	(0.250)	(0.252)
Father has low education	0.504***	0.523***	0.547**	0.722	0.713	0.712
	(0.122)	(0.129)	(0.136)	(0.191)	(0.190)	(0.191)
Lowest decile income		1.539	1.847		0.210	0.212
		(1.041)	(1.263)		(0.254)	(0.257)
Attended a private school		0.769	0.779		1.278	1.277
		(0.257)	(0.262)		(0.412)	(0.411)
Father (bio or step) not present		0.321***	0.294***		0.805	0.799
		(0.117)	(0.109)		(0.259)	(0.258)
Ethnicity						
Reference: Non-black, non-Hispanic						
Hispanic	0.692	0.748	0.935	0.718	0.733	0.726
	(0.379)	(0.415)	(0.534)	(0.419)	(0.432)	(0.432)
Black	1.068	1.209	2.093*	0.713	0.846	0.792
	(0.413)	(0.486)	(0.936)	(0.319)	(0.384)	(0.396)
Standardized AFQT			1.729***			0.958
			(0.309)			(0.171)
Age	4.544	4.884	2.261	1.385	1.750	1.763
	(8.707)	(9.505)	(4.507)	(2.825)	(3.604)	(3.650)
Age^2	0.983	0.982	0.991	0.998	0.995	0.995
	(0.022)	(0.022)	(0.023)	(0.024)	(0.024)	(0.024)
Observations	544					

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C2: Field of study regressions with health and biological sciences included as a separate category

I ran separate logistic regressions with an additional subject group; health and biological sciences, to test the hypothesis that less advantaged US women were sorting into female dominated STEM subjects, rather than studying gender atypical subjects. Results shown in tables 6 and 7 indicate there is no evidence to suggest this is the case. Coefficients on parent's education in predicting study of male dominated STEM subjects remain high. The reduction in sample sizes has a negative impact on interpretability of results, particularly because samples of US men who had studied health and biological sciences were very small. For these reasons, this subject breakdown was not included in main specifications.

Table C2.1: Multinomial logistic regression predicting field of study for women (BCS70)

Base subject: OSSAH	STEM				Health				LEM			
	1	2	3	4	1	2	3	4	1	2	3	4
Social background												
Low parental education	0.951	0.910	0.957	1.032	0.934	0.911	0.960	1.025	0.995	1.004	0.992	0.969
	(0.212)	(0.206)	(0.221)	(0.243)	(0.202)	(0.202)	(0.219)	(0.237)	(0.180)	(0.187)	(0.188)	(0.184)
FSM		2.530*	2.749*	2.730*		1.466	1.602	1.598		2.749**	2.691**	2.709**
		(1.406)	(1.554)	(1.548)		(0.904)	(0.996)	(1.000)		(1.245)	(1.225)	(1.233)
Independent School		0.556	0.543	0.446		0.545	0.530	0.446		1.278	1.285	1.377
		(0.301)	(0.294)	(0.242)		(0.264)	(0.257)	(0.220)		(0.413)	(0.414)	(0.456)
Father (bio or step) not present		0.690	0.697	0.725		0.723	0.735	0.757		0.649	0.649	0.638
		(0.273)	(0.277)	(0.287)		(0.279)	(0.285)	(0.293)		(0.207)	(0.207)	(0.205)
BME	1.479	1.288	1.454	1.474	1.066	1.009	1.151	1.170	1.730	1.541	1.498	1.491
	(1.011)	(0.901)	(1.034)	(1.062)	(0.761)	(0.721)	(0.832)	(0.856)	(0.934)	(0.846)	(0.834)	(0.824)
Cognitive ability (Mean of age 5 and 10 scores)			1.170	1.132			1.186	1.153			0.965	0.974
			(0.166)	(0.162)			(0.162)	(0.160)			(0.106)	(0.108)
Prestige of university												
Mean university acceptance scores				1.004***				1.003**				0.999
				(0.001)				(0.001)				(0.001)
Observations	1054											

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C2.2: Multinomial logistic regression predicting field of study for women (NLSY79)

Base subject: OSSAH	STEM			Health & biological sciences			LEM		
	1	2	3	1	2	3	1	2	3
Social background									
Low parental education	2.081***	2.380***	2.785***	1.762**	1.719**	1.712**	1.682**	1.674**	1.852***
	(0.570)	(0.671)	(0.810)	(0.456)	(0.457)	(0.463)	(0.341)	(0.347)	(0.394)
Lowest decile income		0.373	0.380		1.373	1.370		0.551	0.555
		(0.365)	(0.372)		(0.812)	(0.813)		(0.310)	(0.313)
Attended a private school		3.188***	3.158***		1.059	1.056		1.055	1.052
		(1.175)	(1.174)		(0.485)	(0.485)		(0.390)	(0.391)
Father (bio or step) not present		0.984	1.128		1.229	1.224		1.855**	2.051**
		(0.438)	(0.508)		(0.457)	(0.461)		(0.515)	(0.579)
Ethnicity									
Reference: Non-black, non-Hispanic									
Hispanic	1.077	1.111	1.617	2.016	2.024	2.013	1.593	1.469	1.873
	(0.849)	(0.889)	(1.318)	(1.315)	(1.331)	(1.349)	(0.911)	(0.848)	(1.101)
Black	0.972	1.179	2.193	1.770	1.561	1.573	1.695	1.653	2.456**
	(0.513)	(0.635)	(1.292)	(0.732)	(0.678)	(0.753)	(0.576)	(0.581)	(0.959)
Standardized AFQT			1.837***			1.005			1.470**
			(0.429)			(0.189)			(0.233)
Age	0.00200***	0.00440**	0.00789**	1.125	1.030	1.022	1.667	1.326	1.949
	(0.005)	(0.011)	(0.019)	(2.683)	(2.510)	(2.508)	(3.179)	(2.567)	(3.790)
Age^2	1.074***	1.064**	1.057**	0.998	0.999	0.999	0.993	0.996	0.991
	(0.029)	(0.030)	(0.029)	(0.028)	(0.028)	(0.029)	(0.022)	(0.022)	(0.022)
Observations					626				

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C2.3: Multinomial logistic regression predicting field of study for men (BCS70)

Base subject: OSSAH	STEM				Health				LEM			
	1	2	3	4	1	2	3	4	1	2	3	4
Social background												
Low parental education	1.013	0.942	1.014	1.044	0.828	0.775	0.869	0.932	0.973	0.908	0.893	0.879
	(0.165)	(0.161)	(0.177)	(0.185)	(0.267)	(0.259)	(0.299)	(0.325)	(0.195)	(0.188)	(0.189)	(0.188)
FSM		1.513	1.686	1.654		1.716	2.060	1.984		1.278	1.250	1.260
		(0.759)	(0.851)	(0.834)		(1.452)	(1.749)	(1.688)		(0.743)	(0.728)	(0.735)
Independent School		0.532**	0.508**	0.476**		0.468	0.436	0.368		0.679	0.684	0.705
		(0.170)	(0.164)	(0.155)		(0.324)	(0.304)	(0.262)		(0.242)	(0.244)	(0.255)
Father (bio or step) not present		0.819	0.807	0.827		0.617	0.602	0.639		1.152	1.159	1.145
		(0.257)	(0.255)	(0.264)		(0.408)	(0.400)	(0.429)		(0.394)	(0.396)	(0.391)
BME	2.423*	2.354*	3.069**	3.032**	2.570	2.465	3.725	3.598	1.084	1.077	1.013	1.018
	(1.188)	(1.164)	(1.570)	(1.556)	(2.083)	(2.009)	(3.217)	(3.107)	(0.748)	(0.750)	(0.732)	(0.736)
Cognitive ability (Mean of age 5 and 10 scores)			1.276**	1.257**			1.469*	1.412*			0.951	0.959
			(0.130)	(0.130)			(0.299)	(0.288)			(0.115)	(0.117)
Prestige of university												
Mean university acceptance scores				1.001				1.003				0.999
				(0.001)				(0.002)				(0.001)
Observations						914						

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C2.4: Multinomial logistic regression predicting field of study for men (NLSY79)

Base subject: OSSAH	STEM			Health & biological sciences			LEM		
	1	2	3	1	2	3	1	2	3
Social background									
Low parental education	0.790	0.798	0.885	2.286	2.337	2.077	0.758	0.781	0.776
	(0.180)	(0.185)	(0.209)	(1.513)	(1.594)	(1.460)	(0.178)	(0.186)	(0.187)
Lowest decile income		1.157	1.351		13.42*	8.967		0.278	0.282
		(0.918)	(1.086)		(17.817)	(12.644)		(0.329)	(0.332)
Attended a private school		0.826	0.816		0.301	0.309		1.283	1.285
		(0.289)	(0.288)		(0.530)	(0.545)		(0.408)	(0.409)
Father (bio or step) not present		0.293***	0.269***		0.000000339	0.000000156		0.770	0.773
		(0.111)	(0.104)		(0.000)	(0.000)		(0.235)	(0.236)
Ethnicity									
Reference: Non-black, non-Hispanic									
Hispanic	0.633	0.668	0.802	2.343	2.783	2.186	0.713	0.725	0.715
	(0.382)	(0.410)	(0.501)	(2.570)	(3.127)	(2.599)	(0.413)	(0.423)	(0.420)
Black	0.992	1.239	2.118	1.374	2.000	1.401	0.724	0.835	0.785
	(0.394)	(0.520)	(0.992)	(1.355)	(2.033)	(1.643)	(0.317)	(0.376)	(0.384)
Standardized AFQT			1.704***			0.738			0.951
			(0.324)			(0.371)			(0.167)
Age	8.998	7.895	3.897	15.91	27.91	43.16	1.062	1.204	1.288
	(18.751)	(16.691)	(8.412)	(99.840)	(192.517)	(293.032)	(2.136)	(2.442)	(2.627)
Age^2	0.974	0.976	0.984	0.968	0.962	0.957	1.001	1.000	0.999
	(0.024)	(0.024)	(0.025)	(0.071)	(0.077)	(0.076)	(0.023)	(0.023)	(0.024)
Observations					517				

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix C3: Income regressions predicting the 75^{th} percentile income

It is possible that results differ at different levels of the income distribution, and associations are stronger at the higher end of the distribution. To tests this, quantile regressions at the 75^{th} percentile were run, shown in tables 10 and 11. Whilst overall results are very similar, for women in the US there is a larger difference in the relationship between family background and income in the model controlling only for degree attainment, and the model additionally controlling for subject studied. When accounting for whether US women have a degree, the relationship between parents' education and income is no longer significant. However, when controlling for subject choice, the relationship is again significant (yet still significantly smaller than before accounting for higher educational achievement).

Table C3.1: Differences in log income by family background for employed individuals in the BCS70

	Women			Men		
	Model 1	Model 2	Model3	Model 1	Model 2	Model3
Parents have high education	0.209***	0.116***	0.110***	0.259***	0.169***	0.166***
	(0.0370)	(0.0339)	(0.0333)	(0.0369)	(0.0383)	(0.0363)
Degree		0.509***			0.457***	
		(0.0381)			(0.0401)	
Subject studied						
No degree						
STEM			0.588***			0.489***
			(0.0721)			(0.0502)
LEM			0.764***			0.694***
			(0.0726)			(0.0705)
OSSAH			0.428***			0.279***
			(0.0419)			(0.0583)
BME	0.334***	0.299***	0.287***	0.329***	0.161	0.161
	(0.0957)	(0.0905)	(0.0882)	(0.121)	(0.120)	(0.130)
Cognitive ability	0.124***	0.0815***	0.0802***	0.0948***	0.0667***	0.0674***
	(0.00987)	(0.0101)	(0.00988)	(0.00955)	(0.0108)	(0.0104)
Constant	10.18***	10.06***	10.06***	10.70***	10.63***	10.63***
	(0.0190)	(0.0194)	(0.0190)	(0.0179)	(0.0192)	(0.0190)
N		4,111			3,941	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C3.2: Differences in log earnings by family background for employed individuals in the NLSY79

	Women			Men		
	Model 1	Model 2	Model3	Model 1	Model 2	Model3
Parents have high education	0.139***	0.0626	0.0753*	0.108**	0.0495	0.0399
	(0.0396)	(0.0471)	(0.0437)	(0.0426)	(0.0417)	(0.0404)
Degree		0.373***			0.479***	
		(0.0484)			(0.0529)	
Subject studied						
No degree						
STEM			0.427***			0.516***
			(0.0881)			(0.0600)
LEM			0.492***			0.676**
			(0.0524)			(0.291)
OSSAH			0.270***			0.345***
			(0.0523)			(0.0693)
Non black, non Hispanic						
Hispanic	0.139***	0.0849**	0.0887***	0.0248	0.0105	-0.00552
	(0.0400)	(0.0365)	(0.0336)	(0.0351)	(0.0380)	(0.0390)
Black	0.113***	0.0607	0.0591	-0.0966**	-0.176***	-0.170***
	(0.0392)	(0.0402)	(0.0367)	(0.0393)	(0.0392)	(0.0375)
Cognitive ability	0.264***	0.164***	0.171***	0.258***	0.146***	0.148***
	(0.0222)	(0.0230)	(0.0210)	(0.0198)	(0.0195)	(0.0193)
Age	0.0732	0.0160	0.0300	-0.457	-0.504*	-0.614**
	(0.323)	(0.307)	(0.257)	(0.324)	(0.280)	(0.279)
Age^2	-0.000903	-0.000199	-0.000354	0.00551	0.00615*	0.00740**
	(0.00375)	(0.00358)	(0.00300)	(0.00375)	(0.00324)	(0.00323)
Constant	9.253	10.37	10.06*	20.65***	21.41***	23.84***
	(6.952)	(6.582)	(5.493)	(6.971)	(6.040)	(6.023)
N		2,155			2,335	

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$