Post-compulsory educational aspirations and choices in England

Konstantina Maragkou

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Economics

The University of Sheffield

November 2019

Abstract

This thesis consists of three relevant yet independent empirical studies investigating different questions regarding the determinants of post-compulsory educational aspirations and choices of an English cohort born between 1989-1990 using very detailed survey and administrative data records.

The first study investigates whether the importance of socio-economic background and ability in determining the post-compulsory educational aspirations and choices of young people changed over time by comparing the analytical cohort with an older cohort born in 1970. Educational aspirations and choices are defined as the selection between academic, vocational and no-post compulsory education. The study identifies a decreasing socio-economic effect over time on both aspirations and choices for academic and no post-compulsory education providing evidence that the expansion of academic education has proportionately benefited individuals from all social backgrounds. Further, the study identifies a decreased participation in vocational education which did not arise from falling aspirations but because of rising aspirations and actual participation in post-compulsory academic education.

The second study investigates whether the educational aspirations of secondary school students are influenced by their school peers. Peer effects on individuals' intentions to stay in education are found to be significant for boys but not for girls. Conditional on their plans to remain in post-compulsory education, peers' ability and aspirations to follow an academic rather than a vocational education pathway, have a positive and significant effect on individuals' aspirations to follow an academic route. The study also finds evidence that the provision of information, advice and guidance by schools or external agencies can serve to mitigate peer effects. Finally, individuals with higher ability peers are less likely to have changed their educational aspirations between Year 9 and Year 11 of schooling.

The third study uses detailed administrative records for the whole population of the analytical cohort to investigate the impact of students' socio-economic background on their academic match in 16-19 post-compulsory education. Academic match would occur when students are matched to post-compulsory qualifications studied by similarly attaining peers. Disadvantaged students are found to be more likely to be exposed to academic undermatch compared to their more advantaged peers. The phenomenon is apparent even between

students within the same school. The study also identifies that undermatched students are more likely to be found in disadvantaged schools with lower proportions of high achieving students and higher proportions of ethnic minority and disadvantaged students. In addition, the results indicate that significant masses of undermatched students are more likely to be found in rural districts with higher rates of youth unemployment and higher proportions of poorly educated residents. Finally, the study demonstrates that academic assortative matching has a positive relationship with labour market income returns, at least at early ages.

Overall, the findings of this thesis establish that educational aspirations and choices are influenced by background factors in addition to individual ability and that they are, to a large extent, socially graded. The implications drawn from this research should be important for every policy maker, social scientist, teacher and parent interested in social mobility and equality of opportunity.

Acknowledgements

It takes a village to complete a doctoral dissertation. Endless effort trying to make sense of the data and writing the text over and over. During the past four years, I realised just how vast my support network is. Coming to the end of this long journey, I would like to take this as an opportunity to express my sincere gratitude to all those who were by my side.

First and foremost, I am deeply indebted to my supervisor, ***Steven McIntosh***, for his invaluable guidance and thoughtful insights, but also for his encouragement, unwavering support, patience and deep understanding during these four years. I would also like to express my utmost sincere gratitude to my second supervisor, ***Andrew P Dickerson***, for his meticulous comments, for keeping me motivated with his never-ending excitement about economics and for teaching me how ideas develop. ***Steve*** and ***Andy***, I owe my knowledge in the Economics of Education to you. It has been an honour and a privilege to have been your student. I would also like to thank you for the invaluable support in helping me secure a research position after completing my PhD.

In addition, I would like to thank ***Matt Dickson*** and ***Mark Bryan*** for helpful discussion and comments at my thesis defence. Many thanks to ***Sandra McNally*** for her useful comments and feedback as well as all the other CVER members I have had the pleasure of meeting. The financial support received from the Department of Education, through CVER, in the form of a PhD studentship is also gratefully acknowledged.

I am also grateful to all the academic and support staff in the Department of Economics for all their help and continual support throughout this journey. I am particularly thankful to ***Arne Risa Hole*** for his useful guidance in coding, ***Gurleen Popli*** who gave me the opportunity to further my research experience and ***Panos Nanos*** for his encouragement and career advice. I extend my thanks to all my fellows, both past and present, whom made this journey particularly pleasant. I am especially grateful to ***Christiana Anaxagorou*** for always reminding me that there is a light at the end of the tunnel, ***Bertha Rohenkohl*** for our interesting conversations and ***Maria Petrillo***, who I only met in my final year but provided me with invaluable support and heartfelt friendship.

I am always grateful to my family; my admirable mother ***Elena***, my beloved father ***Kiriakos*** and my sweetheart brother ***Christos***, for always believing in me and for keeping me grounded with their unconditional love and unfaltering support. They walked the way

with me sharing the peaks as well as the valleys.

Last but by no means least, I am eternally grateful to all my wonderful friends and especially to my beloved girls, ***Christina, Maria, Vasilia and Aphrodite*** for always being there for me, picking up my pieces in my most troublesome days. I would also like to express my heartfelt gratitude to my favourite person, ***Danos***, for making the end of this journey so much easier. None of what I have in the following pages would have been possible without all of you.

Notes and Disclaimers

I hereby declare that the work in this thesis was carried out in accordance with the Regulations of the University of Sheffield. I certify that this thesis is solely my own work, unless clearly indicated otherwise. This thesis forms part of the work of the Centre for Vocational Education Research. CVER is an independent research centre funded by the UK Department for Education (DfE). For more details on the Centre, go to cver.lse.ac.uk. Any views expressed in this thesis are mine and do not represent the views of the DfE.

Chapter 2 and Chapter 3 of this thesis are based on data collected from the Longitudinal Study of Young People in England (LSYPE) which has been produced by the Department for Education (DfE) and supplied by the Secure Data Service (SDS) at the UK Data Archive. The data are Crown Copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of these data does not imply the endorsement of the data owner or the UK Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. Chapter 4 of this thesis uses linked administrative data from schools and tax authorities in England. The data have been kindly matched and provided anonymised by officials at the DfE.

An edited version of Chapter 3 has been jointly co-authored with my supervisors, Steven McIntosh and Andy Dickerson and has been published in CVER's Discussion Paper Series (CVER DP 017).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivations and aims

1.1.1 Background and motivation

Inconsistencies in economic development are extensive and in many cases unexplained. Social scientists in their effort to unwind this puzzle bring education in the frontline as empirical evidence suggests that policy interventions in education are significant in promoting social mobility and equality of opportunity for all.

Education, through its impact on economic growth and productivity, is considered as a worldwide solution to some of the most severe problems of the economy such as unemployment, poverty and inequality. Myriads of books and scholarly articles have been published in the field of education, mainly trying to inform how to have more and better-quality education and how to provide equality of educational opportunity. "The *Economics of Education* is about how education is produced, who gets more -or less- education and the economic impact of education on individuals, firms and society as a whole" (Machin and Vignoles, 2018, p. ii).

Countries, in general, tend to invest in education in order to support individuals to obtain skills and knowledge as a form of human capital envisaging to achieve economic value, which will lead to economic growth and development. It was Becker (1964) who first suggested this theory, introducing a framework which analyses why counties and individuals are investing in education like they invest in physical capital. The resulting human capital theory is still forming the basis of most research in the field today. While the important role of education in enhancing human capital in an increasingly complex society and rapid technological change environment is widely acknowledged, the limited public resources to be spent on the educational system require allocations in the most efficient way.

Education economists are very much concerned with issues related to the determinants of educational outcomes including the impact of students' cognitive and non-cognitive skills, school resources and teacher quality. Also, there is a great concern for the demand

for education, the contribution of education to economic growth, the measurements of educational costs and expenditures, the balance between the different types and levels of education and finally, the financial and planning problems arising through its various methods of provision (Woodhall, 1967).

The main role of the education system should be to let the different levels and types of education to be fairly sorted among individuals of differing skills and abilities, independently of socio-economic background, in order for them to be able to look for a job that will provide the maximum returns to the labour market and ensure that the most capable people are fairly sorted in to the right jobs and in the right amounts (Dearden et al., 2009). Recent evidence from Britain shows that social mobility has stagnated since 2014 at all stages from birth to work suggesting that an individual's occupation and income remain, to a great extent, tied to where they started in life (Social Mobility Commission, 2019).

The existence of such inequalities is highly policy relevant as they suggest that the idealised version of education as a social mobility promoter has so far failed to provide equal opportunities to the vast majority of disadvantaged students. Jenkins et al. (2017), in a Sutton Trust's report, suggested that encouraging the opportunity for talent to be recognised across society and providing opportunity for development then this will in turn improve the economy by raising both productivity and Gross Domestic Product (GDP).

Inequality of educational opportunity is considered one of the major barriers to promoting social mobility. There is plenty of evidence that suggests that policies targeted on enhancing social mobility, such as raising equality in university participation or the quality of primary education, can provide to highly able but disadvantaged students the opportunity to generate greater value in the economy (Jenkins et al., 2017).

The issue of vocational education is also a significant concern of the education system. Many social scientists discussed that as a consequence of the expansion of academic education, the vocational education system is forced to suffer a damage. The issue of vocational education is highly policy relevant especially in England, where empirical evidence reports very low returns to vocational qualifications especially when compared to the returns to academic qualifications (McIntosh, 2006, Dearden et al., 2009, McIntosh and Morris, 2016, 2018). McIntosh and Morris (2016) found that individuals holding vocational qualifications at Level 2 and below receive no positive and statistically significant labour market returns even when compared to individuals who hold no qualifications at all. This is not so surprising.

A major reason that employers hold vocational qualifications in lower regard is because the students choosing to study for vocational qualifications are usually those who have the poorest educational attainments (Machin and Vignoles, 2018). Education acts, at least partially, as a screening device, and recent empirical studies suggest that choosing to follow the vocational rather than the academic route is usually a signal of lower cognitive ability (Machin and Vignoles, 2018).

In England, there are additional concerns regarding the vocational education system. While for the students undertaking A levels (which are the main academic qualifications offered for 16-19 year olds) the route is relatively well-known, for the other 50% of students studying mainly vocational qualifications, the system is much more complex and the courses available are much more diverse and not as easy to understand where they lead to (Wolf, 2011, Hupkau et al., 2017). This is an outcome of the proliferation of vocational qualifications which has led to a system that is little understood not only by students but by employers as well. If the skills acquired from the completion of a particular vocational qualification are not clear to employers then it is not surprising that some qualifications have very little economic value (Machin and Vignoles, 2018).

"It is vital that young people have the choice to shape their own lives. This means not only ensuring they get better qualifications and are equipped with what they need to succeed. But it also means making sure they have an informed choice to take up an apprenticeship rather than taking a degree, to find a job which is fulfilling and the choice to stay where they grew up rather than moving away"(Social Mobility Commission, 2019, p. 3). This was the main summary of the Social Mobility Commission (2019)'s report on social mobility in England raising concerns about the importance of informed educational choice. With the aim to inform public policy with regards to these concerns, this doctoral thesis explores the determinants of students' aspirations and choices for post-compulsory academic and vocational education in the context of the English schooling system. In the following section, I introduce the aim of this thesis, delineate the research questions addressed and discuss its contribution in the *Economics of Education*.

1.1.2 Aims, research questions and contribution of this thesis

This thesis consists of three related, yet independent, empirical studies exploring the determinants of post-compulsory educational aspirations and choices of young people in

England. As discussed in the literature, educational *choices* reflect young people's final decisions. Concerning educational aspirations, in the current literature there is no consensus with regards to the definition of the term and especially with regards to what distinguishes aspirations from expectations. Some researchers use both concepts with the same meaning while others define aspirations as 'idealistic hopes for the future' while expectations as 'meaningful realistic plans'.

This thesis adopts the definition of educational *aspirations* as the information used to construct the educational aspirations variables are derived from questions were the respondents indicate their plans and hopes for post-compulsory education (Khattab, 2015). It is often discussed that educational choices are more likely to be directly affected by socio-economic realities compared to educational aspirations which is also what is likely to distinguish them from educational expectations that reflect what realistically one expects to achieve. In this thesis educational aspirations are defined as 'what an individual hopes will happen in the future' while educational choices are defined as 'an individual's final decisions' and are measured in terms of achieved qualifications.

There is a common motive uniting all the chapters in this thesis, and that is their education policy design relevance. Each chapter presents and discusses robust empirical evidence from large-scale data sources to inform on the posed questions.

Building on the theoretical and empirical literature on this field, the following research questions are addressed by this thesis:

Chapter 2 - Study 1:

1. Conditional on cognitive ability, is the importance of socio-economic background in influencing educational aspirations and choices becoming stronger over time?
2. Conditional on socio-economic background, is the importance of cognitive ability in determining educational aspirations and choices becoming weaker over time?
3. Is it only the actual educational choices that are restricted due to socio-economic background and ability or are individuals' aspirations also dominated by these effects?

Chapter 3 - Study 2:

1. What is the causal effect of secondary school peers' ability and educational aspirations on individuals' preliminary (Year 9, at age 14) and later (Year 11, at age 16)

educational participation aspirations?

2. Do peer effects on educational aspirations vary by gender?

3. Does the provision of information, advice and guidance (IAG) reduce peer influence on educational aspirations?

4. Does peer quality affect the probability that individuals will change their aspirations between Year 9 and Year 11 of compulsory schooling?

Chapter 4 - Study 3:

1. Does socio-economic background affect the academic match (studying for qualifications that are achieved by similarly attaining peers) of students in 16-19 post-compulsory education?

2. Are local area characteristics and school composition characteristics influencing the academic match of students in 16-19 post-compulsory education?

3. Is there a relationship between academic assortative matching in 16-19 education and labour market income returns at age 25?

This thesis contributes to the Economics of Education literature in three important ways. First, it provides novel evidence on the impact of socio-economic background, individual ability and school environment (including school peers) in influencing the educational aspirations and choices of students at the end of compulsory schooling. Such decisions are crucial for students' academic progression and labour market success. The thesis also provides a quantitative contribution to the existing literature through making use of very detailed datasets with exceptionally rich information on students' achievements making it possible to rule out -or to clearly identify- the potential role of individual ability in influencing such educational decisions. Finally, the thesis makes important methodological contributions related to the estimation of causal outcomes using instrumental variables and fixed effects as well as unique analytical approaches, providing an inclusive within and between cohorts comparison of aspirations and choices in the second chapter and deriving a new continuous standardised index indicating 'individual ability-qualification selectivity' match in the fourth chapter.

All three empirical studies presented in this thesis are within the context of the English schooling system. To provide an institutional framework the following section provides

relevant information concerning the structure of the compulsory and post-compulsory English education system.

1.2 Institutional background

Compulsory education

In England, the compulsory school curriculum is organised in blocks of years called Key Stages (KS) and it is the same for all students until Year 11 (age 16) but then gives way to a stratified system where the students have to select between an academic or a vocational route or, prior to recent policy allocations, to leave education and directly enter the labour market. From September 2012, the compulsory participation age rose to 17 and from September 2014, this was raised to the age of 18; but students still have to decide whether they will follow an academic or a vocational route. The cohorts that have been analysed in this thesis were not affected by this policy change, meaning that compulsory schooling for them terminated at age 16.

The students in compulsory education at the end of KS are assessed in standard national tests and progress through the phases is measured in terms of KS Levels. In the primary phase, students enter school at age 4-5 in the Foundation Stage, then move to KS1 at age 5-7. At the age of 7-8 students move to KS2 and sometimes, but not usually, with a change of school. At the end of KS2, at age 10-11, children leave the primary phase and go on to the secondary school of their choice, where they progress through KS3 at age 11-13. At both KS2 and KS3 students are assessed in three core modules, Maths, Science and English, and their attainment is recorded in terms of the test scores achieved, which are externally marked as part of the national programme of National Curiculumn assessment. Since 2009 the externally marked exams at KS3 have been replaced with teacher assessments. Finally, the students progress through KS4, at age 15-16, when they take the General Certificate of Secondary Education (GCSE) which used to coincide with the end of compulsory schooling. GCSEs are compulsory high-stake public examinations taken by all school students usually in 9 or 10 subjects.

Post-compulsory education

In the first phase of post-compulsory education, at KS5, the options available to the students staying in education concerning where to go and what to study are very broad.

The students can choose to go to a sixth form in their school (if there is a sixth form in their school) or in another school and study for academic qualifications (usually A levels), they can also go to a sixth form college and finally, they can go to a college of Further Education were they can study mostly vocational qualifications. Also, as mentioned above, while for the students who study for A levels the route is relatively well-known, for the remaining students who undertake the vocational route, the available options are much more diverse and not as easy to understand where they lead to (Hupkau et al., 2017). Finally, the students in post-compulsory education have the option to continue in Higher Education which can be considered the optional final stage of 'formal learning' and is achieved through academic degrees or advanced vocational qualifications.

Table 1.1 summarises the English schooling system, providing information on the equivalent achievement in terms of school years, the age of the students, the duration of the course, and the qualification acquired in each KS.

Table 1.1: English educational system

Key Stage (school year)	**Age**	**Duration**	**Qualification Acquired**
Compulsory education			
Primary education			
KS1 (1-2)	6-7	2 years	KS1 SATS
KS2 (3-6)	8-11	4 years	KS2 SATS
Secondary education			
KS3 (7-9)	12-14	3 years	KS3 SATS
KS4 (10-11)	15-16	2 years	GCSEs
Post-compulsory education			
KS5 (12-13)	17-18	2 years +	AS/A Levels (academic), NVQs/ National Diplomas in vocational routes.
Higher Education	18+	3 years+	Degree

Description of the English educational system divided by Key Stages which correspond to different school years.

Additional information regarding the organization of the compulsory educational system and allocation of students in schools and classes

The Local Educational Authorities (LEAs) are responsible for organising the individuals' admission policies for primary and secondary schools. Following the Education Reform Act (ERA) of 1988 and the introduction of the current National Curriculum, admission to both primary and secondary schools is guided by the principle of parental choice and all students can apply to a number of different schools and attend any under-subscribed school regardless of where they live. Parental choice of school is informed first, formally, by reports made on each school by a government agency (Ofsted) that makes in-depth site

visits to each school and from the summary statistics on each school's performance and second, informally, through friends and general reputation.

This system replaced the previous assignment of children to schools primarily on the basis of residence and the allocation of central government funding to schools by the LEAs. Still, though, there are various other criteria which are used by oversubscribed schools to prioritise applicants which do not involve selection by ability. These criteria usually include to prioritise students with special educational needs, students with siblings in the school and students who live closest. Also the new school system devolved the funding of schools to a more local level provided by central government. As explained by Glennerster (1991), the intention of the new education system was that average funding and parental choice would increase competition between schools for students which would in turn improve educational attainment.

Most households can choose between more than one secondary school from where they live and on average students of the same age who live in the same Output Area (OA)[1] attend two to three different secondary schools every year (Gibbons et al., 2013). Further, the English school system can be described as inflexible in that the school size cannot be quickly increased or reduced (Burgess et al., 2004). A typical English secondary school is attended by students living in around 60 different OAs, meaning that students come from differing family backgrounds. Further, in secondary schools, there is not a unique class, the students are grouped with different peers for different subjects, and therefore they tend to interact with most of the students attending the same school.

State schools cannot select students on the basis of their ability although there are some schools, like the voluntary-aided and foundation schools and especially the grammar schools, which are using selection criteria based on aptitude and ability (West and Hind, 2003). Finally, regarding the organisation of teaching and class formation within schools, although students are grouped with different peers for different subjects it is often discussed that there is a subject-specific allocation of students according to their ability level. Specifically, secondary school students are initially taught in mixed-ability groups for an 'observation and acclimatisation period' which usually lasts around one academic year, and then are eventually educated in different groups for different subjects according to their performance in each subject (Lavy et al., 2012). This is mostly true for GCSE subjects as students are

[1] An OA is a geographic neighbourhood in England comprising an average of 125 households or 1, 500 individuals with 5 students of the same age group

often assigned to classes based on prior achievement as well as academic capacity based on teacher's assessments (Atkinson et al., 2008).

1.3 Structure and content of this thesis

This thesis consists of three separate empirical studies presented in Chapters 2, 3 and 4. Each of these studies utilise student-level data and adopt econometric techniques to analyse influences upon student aspirations and choices in post-compulsory education in England. Chapter 5 concludes this thesis. The four chapters constituting the rest of this thesis are briefly discussed below.

1.3.1 Brief overview of Chapter 2

The first study uses unique longitudinal data from two English cohorts born in 1970, using the British Cohort Study (BCS), and 1989-1990, using the Longitudinal Study of Young People in England (LSYPE), to investigate whether the importance of socio-economic background and ability in determining the post-compulsory educational aspirations and choices of young people changed over time. Educational aspirations outline individuals' future educational plans while choices define the actual decisions that have been taken and are defined as the selection between academic, vocational and no post-compulsory education. Adopting multinomial logistic techniques and a within-and-between cohorts comparison framework, the study evaluates the conditional effects of socio-economic background and ability as well as the differences of these effects between aspirations and choices.

The study, conditioning on individuals' cognitive ability, identifies a decreasing socio-economic effect on both aspirations and choices for academic and no post-compulsory education and provides evidence that the expansion of academic education has proportionately benefited individuals from all social backgrounds. Further, the study identifies a decreasing participation in vocational education which did not arise from falling aspirations but because of rising aspirations and actual participation in academic education. The findings of this study suggest that it is the value of vocational education that needs to be recognised in order to make young individuals aspire to follow it.

1.3.2 Brief overview of Chapter 3

The second study uses the LSYPE to investigate whether the educational aspirations of secondary school students in Year 9 (at age 14) and in Year 11 (at age 16) are influenced by their school peers. Educational aspirations are defined as plans to stay in education after completing compulsory schooling and, conditional on staying, intentions to follow an academic rather than a vocational post-compulsory pathway.

In order to overcome the endogeneity and selection biases associated with peer effects, the study adopts an identification strategy based on 'peers-of-peers'. Specifically, each individual's secondary school peers are instrumented with their primary school peers who did not attend the same primary or secondary school as the individual. These peers-of-peers will have affected the secondary school peers through attendance at the same primary school, but have likely never met the individual and therefore will not have had any direct effect on the individual's aspirations. The study assesses peer effects in two different ways: through peers' ability and peers' aspirations.

Peer effects on individuals' intentions to stay in education are found to be significant for boys but not for girls. Conditional on their plans to remain in post-compulsory education, peers' ability and aspirations to follow an academic rather than a vocational education pathway, have a positive and significant effect on individuals' aspirations to follow an academic route. The study also finds evidence that the provision of IAG by schools or external agencies can serve to mitigate peer effects. Finally, individuals with higher ability peers are less likely to have changed their educational aspirations between Year 9 and Year 11 of schooling.

The results have implications for allocations of students across schools. Even in a mostly comprehensive education system as in England, with no selection by schools on ability, there are still large differences in student intakes across schools. The findings of this chapter suggest that the provision of more IAG would benefit those young individuals whose aspirations are influenced by their secondary school peers.

1.3.3 Brief overview of Chapter 4

The third study investigates the impact of students' socio-economic background on their academic match in upper secondary post-compulsory education using detailed

administrative records from schools, colleges and tax authorities in England. Students are academically matched when studying for qualifications that are achieved by similarly attaining peers. Academic match is identified using a continuous measure of student-qualification match which identifies undermatched, matched and overmatched students based on the distance of each student's attainment in age 16 high-stake examinations from the median attainment of students studying their chosen academic or vocational qualification. The study is novel in that it is the first to examine academic match at the upper-secondary level and it is the first to take vocational qualifications into consideration.

Disadvantaged students are found to be more likely to be exposed to academic undermatch compared to their more advantaged peers. The phenomenon is still apparent even when comparing students within the same school. The study also identifies that undermatched students are more likely to be found in schools with lower proportions of high achieving students and higher proportions of ethnic minority and disadvantaged students. In addition, the study suggests that significant masses of undermatched students are more likely to be found in rural districts with higher rates of youth unemployment and higher proportions of poorly educated residents. Finally, the study demonstrates that academic assortative matching has a positive relationship with labour market income returns, at least at early ages.

There are important policy implications to be drawn from these findings. Policy-makers interested in social mobility should be focusing more on providing students with IAG related to the available 16-19 education courses that are suitable to each student's ability credentials and future educational and occupational aspirations.

1.3.4 Brief overview of Chapter 5

This final chapter concludes this thesis, provides areas of future research and discusses some policy implications derived from the main findings. The results of this thesis show that young people's educational decisions are socially graded, with students from disadvantaged backgrounds consistently found to aspire to and choose different post-compulsory educational routes compared to their more advantaged peers. This thesis shows that this social bias has been significantly reduced between the years, yet not disappeared. The post-compulsory educational choices of disadvantaged youth are found to be undermatched to their academic credentials causing severe costs on their labour market income returns. Further, this

thesis establishes the importance of secondary school peers in influencing the educational aspirations of young people and that the provision of IAG can serve to mitigate the importance of peers in influencing these aspirations.

Overall, the findings of this thesis use robust empirical evidence with the aim to inform policy makers about the determinants of educational aspirations and choices and guide them to develop the English educational system; motivating towards providing more information about post-compulsory education options, as a step closer to social mobility and equality of opportunity in education and the labour market.

Chapter 2

The changing influence of socio-economic background and ability on post-compulsory educational aspirations and choices

2.1 Introduction

2.1.1 Background and motivation

Improving the quality of the educational system has been a key area of research, especially for the last few decades where educational participation has experienced an almost unprecedented expansion (Galindo-Rueda and Vignoles, 2005b). In response to this expansion, a considerable literature exists that examines the possible influences and outcomes of the increasing educational attainment as well as the related issue of educational inequality. Hupkau et al. (2017) indicated that the students who follow academic education tend to have much higher prior attainment and are much less likely to come from a disadvantaged background than the average student and that on the other hand, progression routes for students who do not undertake academic qualifications are not as well-known or preferred.

Equality of opportunity concerns have taken the form of improving the educational attainment of high ability students from disadvantaged backgrounds (Galindo-Rueda and Vignoles, 2005b). Plenty of evidence suggested that better educated parents or parents from a higher social class provide a 'better' environment for their children which has form the basis of several policy interventions in education such as the introduction of a standardised national curriculum for all students aged 7-16, school choice and the reform of the vocational education system. Also, it is widely believed that while higher maternal and paternal education has more or less similar impact on family wealth, the external effects related with education are greater in magnitude for maternal rather than paternal education because mothers are usually those providing care within the household, especially in early years (Chevalier et al., 2013). In addition, the skills and aspirations that children

develop between young ages have been proven to form a central component in influencing young people's future educational participation and subsequently their labour market success. Erberber et al. (2015) identified that educational aspirations are the strongest and most consistent predictor of academic success of disadvantaged students.

All the issues discussed above provide crucial evidence of educational inequality and make an essential argument for funding young people's education, especially of those from less advantaged socio-economic backgrounds. While intergenerational correlations in education are not doubted, any suggested policy implications should be highly dependent on "the characteristics of the intergenerational transmission mechanisms and the extent to which the correlation is causal" (Chevalier et al., 2013, p. 1).

From a research perspective therefore, the main empirical questions are whether educational expansion contributed towards family background becoming less important in determining educational aspirations and choices, and, consequently, whether individual ability is becoming a more important factor in predicting these outcomes. This chapter considers this crucial policy issue in the context of the English education system, which makes for an interesting case study because it has been exposed to some major policy changes in the post-war period, and is also experiencing a significant expansion in educational participation over the last 50 years (Galindo-Rueda and Vignoles, 2005b). Specifically, England the past years has been undergoing a massive expansion of academic education followed by a decline in participation in vocational education.

2.1.2 Research question

The aim of this chapter is to investigate the changing influence of socio-economic background and cognitive ability on post-compulsory educational aspirations and choices using two cohorts of young individuals, the older born in 1970 and the more recent born 20 years later, between 1989-1990. Educational aspirations are the goals that individuals have for their future as expressed close to the end of compulsory schooling while educational choices are their actual decisions in terms of what qualifications they achieved at a later age. Educational aspirations and choices are defined as the selection between academic, vocational and no-post compulsory education.

The chapter does not aim to estimate the causal effect of prior ability and socio-economic background on educational aspirations and choices but rather, follows a strongly comparative

approach to examine, *between* cohorts, the change in the importance of these inputs in influencing educational aspirations and choices. The comparison of the socio-economic and ability gap between the two cohorts is particularly interesting given that the older cohort completed compulsory schooling in 1986 before the 1988 Education Reform Act (ERA) and the expansion of Higher Education which is highly relevant to the choices facing 16-year olds and therefore provides important context. Further, the chapter aims to investigate the difference in the effect of these inputs between aspirations and choices *within* the same cohort. Finally, the chapter aims to explore whether the influence of socio-economic background varies between maternal and paternal education and occupation.

In summary, this chapter aims to discover whether educational expansion changed the importance of family background in influencing individuals' aspirations and choices for academic, vocational and no post-compulsory education and whether the influence of ability, conditional on the socio-economic background, became less important in influencing educational aspirations and choices. In addition, the differences between educational aspirations and choices have been examined and through an inclusive within cohorts analysis, it is determined whether it is only the actual choices that are influenced by ability and, most importantly, social class or whether individuals' aspirations are also dominated by the socio-economic and ability effects.

2.1.3 Research findings and limitations

The main findings of this research are summarised as follows:

- There is an expansion of academic education and a decline of vocational education which is found not to be arising from falling aspirations to attend vocational education but because of increasing aspirations and actual participation in academic education.
- The socio-economic effect on educational aspirations and choices for academic and no post-compulsory education is significantly decreased between the older and the more recent cohort suggesting that the expansion of academic education has proportionately benefited individuals from all social backgrounds.
- An individual's early ability became a poorer predictor of educational aspirations and choices for academic and no post-compulsory education.
- The socio-economic and ability effect on vocational aspirations is becoming increasingly

negative from the older to the more recent cohort. For final choices, while the ability effect is becoming more negative, the socio-economic effect starts out negative in the older cohort but becomes less negative in the more recent cohort implying slightly decreasing socio-economic effects.

These findings have their limitations in terms of the older sample suffering from attrition in the dataset. The chapter overcomes this limitation by applying some structural allocations in some of the variables (explained in detail in Appendix A.5), ensuring that the analytical sample is representative of the total sample. Nevertheless, it should be acknowledged that some findings might be biased due to these allocations. Possible multicollinearity problems which could arise in the estimations due to high correlation between parents' occupation and education have been overcome by estimating the two socio-economic components in separate regression models.

2.1.4 Structure

The remainder of this chapter is set out as follows. In the next section there is a comprehensive review of the relevant literature on education, family background and ability, focusing mainly on British studies. Section 2.3 describes the data. Section 2.4 explains the methodology that has been used to estimate the link between socio-economic background, cognitive skills and educational aspirations and choices. The main empirical results are presented in Section 2.5. Finally, Section 2.6 concludes this chapter.

2.2 Literature Review

2.2.1 The effect of ability and social background on educational aspirations and choices

There is an expanding and controversial literature on the role of education in developing or declining social mobility and equality of opportunity. This chapter relates to the broader literature on the link between education and inequality as well as on the literature focusing on the effects of individuals' ability. De Fraja (2002) supports that the 'ability to benefit from education' is a combination of both family background and innate ability.

There is evidence supporting the idea that parents from higher social classes impact their children's cognitive ability progression by spending more resources and investing more money

on them (Galindo Rueda and Vignoles, 2003). Carneiro et al. (2010) explain that children from disadvantaged backgrounds tend to not only start with less resources but also attend lower quality schools which has a negative effect on their formation and the improvement of their skills and knowledge. Galindo Rueda and Vignoles (2003) argue that although cognitive ability in Britain is a very important determinant of high earnings and labour market success, the evidence supporting that an individual's socio-economic background has become a more important determinant of cognitive development, undermines this argument.

In the empirical literature, parental influence has been proven crucial for the educational outcomes of their children which starts from both the parents' and children's ambitions and aspirations about their future. Croll (2008) using data from the British Household Panel Survey (BHPS), shows evidence of young people from advantaged families being more ambitious, achieving better educationally and having better labour market outcomes.

Schoon and Duckworth (2010) used three cohorts of British and English individuals born in 1958, 1970 and 1989-1990, with data collected from the National Child Development Study (NCDS), the British Cohort Study (BCS) and the Longitudinal Study of Young People in England (LSYPE), to examine whether the educational expectations of individuals to stay or leave full-time education, after completing compulsory participation at age 16, are influenced by their own ability and school motivation, their parents' years of schooling and their parents' expectations, all this in a changing social context. The study identifies that educational expectations in the most recent cohort (LSYPE) have increased significantly and that they are less associated with parental education and prior academic attainment suggesting that expectations for post-compulsory education participation are becoming the norm. Further, the authors identify that socio-economic inequalities in academic attainment are persistent while the gender gap in expectations is expanding, with boys being less ambitious for their future education than girls.

Berrington et al. (2016) using data from the UK Household Longitudinal Study (UKHLS) identified that White boys from disadvantaged backgrounds have the lowest aspirations for university attendance. Further, the authors mentioned that the socio-economic and ethnic background of individuals act as important mediating factors for parental attitudes towards education, levels of parental engagement with their children's education and the quality of the parent-child relationship as well as individual's educational attainment.

Goodman et al. (2011) support that aspirations for university participation are socially graded, which implies that young people from poorer backgrounds have lower educational aspirations than their more advantaged counterparts and are more likely to display risky behaviours both as teenagers and as adults. Consequently, even if a young person's disadvantaged background restricts them from realising their educational aspirations, the fact that disadvantaged students are more likely to have lower aspirations to start with is what forms the main issue for concern. Guyon and Huillery (2016) used a sample of French teenagers to establish the importance of aspirations in influencing important future school outcomes and also to identify that social inequalities in educational aspirations are not driven by differences in professional aspirations. Further, Gutman and Akerman (2008) find that British girls, from minority ethnic groups and from higher socio-economic backgrounds tend to hold higher aspirations than their peers.

Apart from the limited literature on educational aspirations, there is a considerable literature examining the role of family background on educational participation and outcomes. Chevalier and Lanot (2002) attempted to estimate the relative importance of ability on educational attainments in Britain, comparing the cohort of the BCS, born in 1970 , with the cohort of the NCDS, born in 1958. Using an ordered probit model with dependent variable the age that the pupils had been leaving education (from age 16 to 20) they introduced a methodology which separated familial and financial effects. Family characteristics included parental education, fathers' occupation, number of siblings, indication for the presence of natural parents and ethnicity as well as several neighbourhood composition characteristics. The model controlled for ability using test scores, aiming not to reflect only the natural ability of the child but also the material and emotional support provided by the parents in an attempt to distinguish between the direct and indirect effect of parental income; suggesting that ability is a function of unobserved family background characteristics. Chevalier and Lanot (2002)'s results provide evidence which supports that even though students from disadvantaged backgrounds are less likely to invest in education, a financial transfer would not lead to a significant increase in schooling investment as several effects of family characteristics which affect the development of the child, as measured by the latter's cognitive skills and abilities, dominate the financial constraint effects.

Adding to this literature, Galindo-Rueda and Vignoles (2005b) using the same samples and controlling for ability using the same, or similar, tests estimated a generalised ordered logit model to allow them not to use identical explanatory variables across thresholds.

The dependent variable consisted of the highest academic qualification level achieved and was separated into 5 categories, from no qualifications to degree or above. The results concluded that individuals' cognitive ability was less important in determining educational participation for the most recent cohort, born in 1970, compared to the older cohort, born in 1958, while socio-economic background became a more important determinant. An interesting issue that arose out of these empirical results, contradicting the previous theoretical literature which supported that improved educational opportunities for disadvantaged students would lead to less education inequality (George and Jr., 1996, Fernandez and Rogerson, 1996, 1998), is the radical shift from selective to mixed ability schooling. The statistical results from this study yield an income-driven educational inequality effect which seems to have increased between richer and poorer pupils as family background has been proven to be the major factor indicating educational attainment.

In an attempt to exploit the trends that follow the access to Higher Education in Britain, Egerton and Halsey (1993) followed a different methodology where they attempted to use as a dependent variable not the type of education but the type of institution in which qualifications are gained including universities, polytechnics and colleges. With evidence from the General Household Surveys (GHS) of 1985,1986 and 1987, Egerton and Halsey (1993) used logistic regression techniques in the Generalised Linear Interactive Modelling (GLIM) package. The authors associated attendance at colleges with the acquirement of vocational work-related qualifications and supported that students from a service-class background; including employees in government, private economic and social service and in general any employee whose employment relationship is based on a code of service rather than a labour contract (Scott and Marshall, 2009), are more likely to have achieved their qualification in a university compared to those from an intermediate or working-class background whose degrees where more likely to be undertaken in polytechnics or colleges indicating that more privileged individuals tend to dominate more prestigious institutions. The empirical evidence from this study suggests that since the beginning of the current century there has been a period of considerable expansion of education where there has been a noticeable decrease in gender inequalities but there was no reduction in relative socio-economic inequality.

Jerrim and Vignoles (2013) further analysed the effect of the socio-economic influence on the education of highly able children from disadvantaged backgrounds using data from the Millennium Cohort Study (MCS). After considering the methodological difficulty named

regression to the mean (RTM) the authors found "no convincing evidence that able but disadvantaged pupils fall behind their more advantaged peers". This evidence supports these results for individuals that are between the age of 3 and 7 and before the end of primary school suggesting no strengthening impact that would cause socio-economic status gaps in children's cognitive achievement to widen. In addition, Dolton and Vignoles (2000) show that poor prior attainment is a more important determinant of university participation among students from disadvantaged backgrounds than the barriers arising at the point of entry into university.

In addition to the parental influence and parental occupation which act as a key structural consequence of their children's future, Ermisch and Francesconi (2002) mentioned that parents' educational attainment is found to be a very strong predictor of their children's educational attainment. With evidence from the first seven waves of the British Household Panel Survey (BHPS), Ermisch and Francesconi (2002) focused on a sample that was born between the period 1974-1981. Using an ordered logit model for educational attainment, a latent dependent variable was included measuring the level of education for each individual in the chosen family. In addition to the strong correlation found between parents' and children's education, the results show evidence of a strong negative association between single-parent families and educational attainment as well as with low income parents or within families with more brothers and sisters. A positive relationship with educational attainment was found for the children whose parents were home-owners and particularly outright owners. It is worth mentioning that Ermisch and Francesconi (2002)'s model does not control for individual ability but only for family background effects.

Further to educational choices, parents' education has an impact on the economic exposure of their children. There is evidence that poorly educated individuals are likely to be relatively more prone to economic vulnerability because they may be more likely to have low-educated parents and low educated peers (Gesthuizen and Scheepers, 2010). Testing the implications of the relative risk aversion hypothesis (RRA) of educational choice which supports that the parents' education is determining their children's educational choices, Davies et al. (2002) find evidence that partly supports this idea suggesting that young people's educational choices are made so as to minimise the risk of ending up with a lower level education than their parents[2].

[2]The theory of Relative Risk Aversion (RRA) suggests that educational decision making is motivated by the individual's desire to avoid downward social class mobility and, furthermore, that this desire is stronger than the desire to pursue upward mobility (Holm and Jaeger, 2008).

In a micro-analysis of Higher Education participation and focusing on two samples collected from Cohort 7 (members aged 18 in 1996) and Cohort 9 (members aged 18 in 2000) of the Youth Cohort Study (YCS), Galindo-Rueda et al. (2004) developed a probit model indicating whether individuals where attending university at age 18. The authors controlled for background characteristics of the cohort members including parents' education, social group and type of school attended. The study, due to lack of available data, used a proxy of prior educational attainment using age 16 (GCSE) and age 18 (A level) test scores. The results of the study show evidence that in the first sample, for a given level of prior achievement, family background did not have an additional impact in the decision of going to university. In the later sample though, that was not the case. The impact of coming from a well-off family when controlling for parents' occupation is positive even when controlling for A levels suggesting an increase in the socio-economic gap in UK Higher Education. A small decrease though had been identified in the impact of parental education as well as a reduction in the impact from attending selective (grammar) schools. It has been noted though, that when not controlling for the type of school the impact of socio-economic background increases suggesting that the school type is an endogenous variable which reflects the pupil's prior achievement. When the authors' analysed models which included finer measures of educational achievements they concluded that the social class effects became small and insignificant.

2.2.2 Economic impact of academic education expansion and vocational education decline

Evidence on the economic impact of and demand for more graduates is ambiguous and contradictory. There are studies which support that a further expansion of Higher Education might lead to a decline in social mobility and that the labour market opportunities for those without degrees or with qualifications below degree level may get worse (Keep and Mayhew, 2004).

Results from various research papers on vocational training reveal that the vocational options and the returns to vocational qualifications vary hugely across different countries (Carneiro et al., 2010). Di Stasio et al. (2015) and Van de Werfhorst (2011) supported that in societies with developed vocational schools, education tends to act more with the human capital approach and logic whereas in industries where they dispute the value of vocational

training and confront it with a more judicious manner, education acts more as a positional good. In addition the literature supports that a lower incidence of over-education is found in countries with strong vocational educational systems. Barone and Van de Werfhorst (2011) justify that in countries with detailed vocational training programs where students are enrolled in both school and the work place, such as happens in the German-speaking countries including Germany, Austria and Switzerland, employers have the chance to select on the basis of productive skills. Acemoglu and Pischke (1999) and Lynch (1993) identified that the countries which have well developed vocational training and apprenticeships systems experience high income returns to vocational education probably because of being more recognised and trusted by employers as well as because of the existence of a competitive market for apprentices. When this is the case, the human capital approach acts as a more relevant mechanism for the effect of education on the labour market as opposed to countries like the US and the UK where the education system is more generally oriented and is aimed at generating competencies among its students and as a result acting as a positional good.

Lynch (1993) pointed out that the vast majority of young people who used to enter vocational training in Britain had school attainments that were well below the level at which they would be accepted for any form of apprenticeship in Germany suggesting that the British apprenticeship system used to be deficient and implied that employers and trade unions need to make an effort to improve for the advantage of young people, firms and the economy as a whole. Before the early 1980s, and particularly before the introduction of the National Vocational Qualifications (NVQs) which were set up in 1986, the UK, indeed, used to experience an undeveloped vocational system where there has been noticed a dramatic shift from apprenticeships. As described by Carneiro et al. (2010) and Dearden et al. (2009) the proliferation of vocational qualifications had lead to a condition where vocational qualifications had little economic value in the British labour market and provided a weakened signal of the abilities that vocational training was providing leading to decreased income returns in any form of vocational qualification with substantial difference from returns to academic qualifications (McIntosh, 2006, Dearden et al., 2009, McIntosh and Morris, 2016).

Using data from the NCDS and the BCS, De Coulon and Greenwood (2015) investigated the effect of the reforms on vocational education occurring in Britain during the 1980s. The study mainly investigated the effects of the removal of the government support for firm-based

training and the introduction of the new vocational qualifications, mainly including General National Vocational Qualifications (GNVQs) and Business and Technology Education Council qualifications (BTECs). These reforms aimed to regulate the shortcoming of apprenticeships in Britain and create higher-status vocational qualifications. The results they came up with suggest that individuals who left compulsory education in 1974, before vocational education reforms, and chose to follow vocational courses earned higher wages than those directly entering the labour market. Following the sample of the BCS though, the results show that individuals who followed vocational training in 1986, after the reforms, obtained a wage penalty over the individuals who directly entered the labour market after leaving compulsory education. The interpretation that De Coulon and Greenwood (2015) gave to this effect is that probably individuals of average ability moved away from vocational training following other educational routes and allowed vocational training to be accessed mainly by lower ability students.

De Coulon and Greenwood (2015)'s analysis is supported by the findings of Payne (2003) who using data from the YCS, which interviewed English youth who were eligible to leave school in 2000-2001, after the reforms in vocational education, attempted to associate vocational qualifications with individuals' characteristics. Using a multinomial logit model with dependent variable academic, vocational and no post-compulsory qualifications and controlling for ability using GCSE scores, she classified that vocational, as opposed to academic, study at ages between 16 and 17 is associated with lower attainment in GCSE exams, higher levels of truancy, less educated parents, rented accommodation, and state rather than private schooling. The factors associated with vocational qualifications rather than having no qualifications included among others good GCSE results, ethnic minority backgrounds, more skilled and better qualified parents and owner-occupied accommodation.

In the modern economy of the UK it has been identified that greater emphasis had been given in expanding academic education which would lead to a desired more meritocratic and socially equal nation by providing opportunities for economic advancement through the expansion of Higher Education provision for students from the lower socio-economic groups. Although the previous findings show a disappointing outcome of vocational training's reforms, it is worth mentioning that since 2004 there has been a huge growth in the number of apprenticeship starts in the UK and particularly since 2010 due to support from successive governments. Evidence from Delebarre Jeanne (2016) suggested a good effort for improvement of the British apprenticeship system as well as an association

of recognition and acceptance of the importance of vocational training. Further, recent evidence from McIntosh and Morris (2018) shows that individuals who completed an apprenticeship experience a significant increase in earnings compared to individuals who started an apprenticeship but did not complete it.

2.2.3 Contribution to the literature

The literature reviewed indicates that there are mixed results and opinions on the effect of socio-economic background and ability on educational aspirations and choices. The various studies mentioned use datasets from different sources and the samples differ in terms of time period and country coverage -although the review is mainly focused on evidence from Britain. Some studies control for ability and others do not leading to results which cannot be directly compared to this chapter. There are some previous analyses which control for ability but not with cognitive skills developed at early ages including only proxy measurements from later school achievements. Further, the majority of previous studies exclude vocational education from the estimation. This chapter makes use of two very rich longitudinal datasets with detailed information on socio-economic background and cognitive skills and adopts a Multinomial Logit Model which allows us to investigate the importance of these effects in influencing educational aspirations and choices while considering all three main pathways that individuals can follow after completing compulsory schooling; including academic, vocational and no post-compulsory education.

Table 2.1 below summarises the methods and findings of some of the key studies in this field. The present chapter follows a similar approach to that of Payne (2003) using different data sets and other measurements of ability. The studies which are summarised below use similar indicators of socio-economic background including parents' occupation, family income, parental education, type of school attended and private or rented accommodation.

In consideration of the key studies summarised below the initial aim of this chapter is to compare the change in the direction and magnitude of the socio-economic and ability effect through the years. There are studies in the past which attempted to do a comparison in different time periods as well. Galindo-Rueda and Vignoles (2005b) specifically used the sample of BCS as their recent sample and compared with the sample of NCDS which is an older one. The interest of this chapter is that the same sample of BCS is compared with a more recent one (LSYPE) adding a further dimension to the previous literature. The

sample of LSYPE completed the compulsory educational level in England in 2006 and had to decide whether to follow post-compulsory education or whether to directly enter the labour market. The results from this sample have been compared with the sample from the BCS which finished compulsory schooling in 1986 and took the same educational decisions 20 years before. Another advantage of the used data is that LSYPE as well as BCS have plenty of information on each member's cognitive skills taken from school attainments or from separate tests that the members had been asked to complete as part of the survey.

The novelty of this chapter is that apart from a comparison *between* the different cohorts, the differences between educational aspirations and choices *within* the sample of the same cohort are investigated through an inclusive, within cohorts analysis. The contribution of this chapter to the existing literature is that through the analysis of datasets rich in information about the cohort members and family background, it does not only identify the magnitude and direction of the socio-economic and ability effect through the years but also identifies whether individuals' socio-economic background is influencing their aspirations by forcing poorer individuals into lower educational aspirations or whether individuals from poorer backgrounds aspire just as much as more well-off individuals but their actual educational choices are restricted because of their family background.

Table 2.1: Summary of key studies

Study	Sample	Estimation	Ability	SES	Qualifications	Results
Ermisch and Francesconi (2002)	BHPS (born between 1974-1981)	Ordered Logit	Does not control for ability	Parent's education, Family income, House tenure	Academic, Vocational, No qualifications	Positive SES effects
Payne (2003)	YCS Cohort 11 (born:1984)	Multinomial Logit	Proxy measurement using GCSE scores	Parents' occupation, House Tenure	A levels, NVQ Level 3, Other qualifications	Positive SES and ability effects
Galindo-Rueda et al. (2004)	YCS Cohort 7 (born:1978) Cohort 9 (born:1982)	Probit	Proxy measurement using GCSE scores	Parents' occupation, Parents' education	Participation in HE	Small and insignificant SES effects, positive ability effects
Galindo-Rueda and Vignoles (2005a)	NCDS (born:1958), BCS (born:1970)	Generalised Ordered Logit	Reading, Maths, General ability scale (age 10-11)	Father's occupation, Parents' age when left schooling	Highest academic level (from no qualifications to HE)	Increasing SES and decreasing ability effects
Schoon and Duckworth (2010)	NCDS (born:1958), BCS (born:1970), LSYPE (born 1989-1990)	Probit	Reading, Maths, General ability scale (age 10-11), KS2 scores (age 11)	Parents' years of schooling, Parents' expectations	Expectations to stay in education after age 16	Decreasing SES and ability effects

2.3 Data

The cohorts that have been used in this chapter consist of young people born in England in two different time periods. The information for the individuals has been collected from two longitudinal studies, namely, the British Cohort Study of 1970 (BCS) and the Longitudinal Study of Young People in England (LSYPE). The principal advantage of the two surveys is that they contain information not only on family background and educational attainment but also measures of cognitive ability prior to the age of 11. The two studies are not identical since the data from the respondents were not acquired at exactly the same ages and the questions were not identically structured but the correspondence of the ages as well as the similarity of the questions being asked allowed robust cohort comparisons to be made.

In addition to structural changes between the two cohorts attrition from the two panels has been also considered. Particularly, I was concerned that sample attrition would be greatest among individuals from less privileged family backgrounds and therefore response bias

at the individual level would underestimate the magnitude of the socio-economic effects. Nonetheless, under some structural allocations to certain variables which are explained in detail below, the analytical sample of both studies remained closely similar to the total sample. Below, all the information is provided about the background of the two studies, an analysis of the attrition in each cohort, descriptive statistics and information for all the variables that have been used.

2.3.1 Description of the Datasets

The British Cohort Study

The data which represent the older sample have been collected from the English population of the BCS. The survey follows the lives of more than 17 000 people born in the UK during one particular week of April 1970. The study is a continuing, multi-disciplinary longitudinal survey following the development of its' participants. Although the original scope of the study had been medical aspects at birth, over the course of cohort members' lives, the BCS has broadened and information has been collected on aspects including physical, educational and social development, and economic circumstances among other factors. The BCS is conducted by the Centre for Longitudinal Studies (CLS) and the information collected is available from the UK Data Archive. Since the birth survey in 1970, there have been eight sweeps of all cohort members at ages 5, 10, 16, 26, 29, 34, 38 and 42 and in addition to the main BCS follow-ups, there have been five BCS special sub-studies where additional data have been collected for samples of cohort members selected for their particular characteristics or circumstances. As mentioned above, an advantage of the BCS is that this specific birth survey contains an unusually wide range of information on respondents' family background, their parents' education levels and their own early cognitive development and their eventual educational attainment.

The data included in this research have been collected from the various sweeps and contain information from questionnaires answered both by the parents and the cohort members. The final analytical sample that has been used for this chapter consists of 6,652 individuals with non missing information on socio-economic background, ability, educational aspirations and choices as well as various other background characteristics.

The Longitudinal Study of Young People in England

The data which represent the more recent cohort have been collected from LSYPE which is a longitudinal study that follows the lives of around 16 000 individuals born between 1 September 1989 and 31 August 1990 in England. The study began in 2004, when the cohort members were aged 13-14 and included a representative sample of young people in Year 9 who attended state and independent schools. The main role of the study is to provide evidence on the key factors affecting educational progress and attainment and the transition following the end of compulsory education. The study apart from education has also collected information about the members' attitudes to school, aspirations for future work and study, use of leisure time, economic circumstances, friends and family life, physical and emotional health and well-being as well as social participation and attitudes. Following the initial survey at age 13-14, the cohort members were visited every year until 2010, when they were aged 19-20. An additional survey took place in 2015 and interviewed the individuals when they were 25 years old. In addition to the young people themselves, parents or guardians were also interviewed in the first four waves of the survey in order to acquire a definite view of the young peoples' households.

Apart from the wide range of information that is provided for the individuals themselves and their family, one major advantage of using LSYPE is that it is the only national longitudinal survey focusing on young people born in the early 1990s, following their pathways through the teenage years and their transitions to adulthood. Another advantage of using LSYPE is that even though the dataset itself does not contain any information on the cohort member's cognitive skills it can be matched to the National Pupil Database (NPD). The NPD is a longitudinal administrative dataset which records all school and college pupils in England throughout their schooling years providing detailed information on their prior test scores and exam results alongside pupil characteristics and school information. This match of the LSYPE data to the NPD is very important for the needs of this chapter as the pupils' KS2 Maths test scores taken at the age of 10-11 were used as indicators of the individuals' cognitive skills.

As with the BCS, the data have been collected from the various sweeps of the survey. The final analytical sample consists of 9,405 individuals with non missing information on socio-economic background, ability, educational aspirations and choices as well as various other background characteristics. The composition of the analytical samples is explained in detail below.

2.3.2 Attrition

Missing data in longitudinal analysis constitutes a major problem for two main reasons. First, missing information leads to the loss of observations and to the reduction of sample size and secondly, non-random attrition would lead to biased estimations. It has been observed that the probability of attrition occurs non-randomly and often it is associated to some observable characteristics such as socio-economic background and education (Mostafa and Wiggins, 2014). If attrition is indeed associated to any observable or unobservable characteristics of the individuals then failing to take it into account would lead to the loss of a particular type of respondents and therefore the sample will no longer be random or representative of the population.

Attrition was much more of a problem for the BCS dataset rather than for the LSYPE dataset. Examples of previous studies which attempted to handle attrition in the BCS include the study of Galindo-Rueda and Vignoles (2005b) who used weights to adjust the distributions of the respondents so that the importance of each cohort member's characteristic is re-weighted according to the importance of the characteristics of those who dropped out and the study of Schoon and Duckworth (2010) who used multiple imputation by chained equations (ICE). Of particular interest is the study of Schoon and Duckworth (2010) who attempted to compare LSYPE with BCS (and NCDS). Schoon and Duckworth (2010) handled attrition by carrying out pathway analysis using the statistical package Mplus 5 but, still, the authors had to acknowledge as a limitation of the study that first, the BCS (and NCDS) is a largely representative study of young people born in Britain while LSYPE is based on young people attending schools in England only and second, that missing data might have affected the validity of the results.

In this chapter the first issue has been handled by keeping only the BCS participants who resident in England. Concerning the second, the chapter does not make use of any imputation methods but instead applies some structural allocations, such as adding missing observation dummies to certain variables which had particularly low response rate, in order to keep more individuals in the analytical sample. Any structural allocations that have been undertaken are explained in detail in the variable sections below. The final analytical sample for both cohorts remained highly representative of the total sample. Table 2.2 presents the attrition details of the two samples. A detailed comparison of the descriptive statistics of the total and analytical sample can be found in Appendix A.7.

Table 2.2: Attrition Details

	BCS	LSYPE
Observations		
Total in cohort	17916	
Total in cohort (age 10 follow-up)	13699	
Total in cohort (English)	12118	
Total sample	12118	15770
Sub-sample: Observations without missing aspirations and choices	7853	11012
Sub-sample: Observations without missing aspirations, choices and ability	6739	10125
Sub-sample: Observations without missing aspirations, choices, ability and control variables	6652	9405
Analytical Sample	**6652**	**9405**

The total BCS cohort are considered the individuals who participated in the age 10 follow-up (in order to be able to obtain their prior ability test scores) and mentioned England as their standard region of residence in the age 16 follow-up (when they were attending secondary school) for comparability purposes with LSYPE. The SES variables (parents' occupation and education) for both cohorts have been treated with an additional category of missing observations.

2.3.3 Variables

Dependent Variables

The aim of this chapter is to identify the change in the influence of socio-economic background and cognitive ability on post-compulsory educational aspirations and choices from the older cohort born in 1970 to the more recent cohort born between 1989-1990. The dependent variables used in the analysis are categorical variables capturing the individuals' aspirations and choices for academic, vocational and no post-compulsory education.

Educational Aspirations

The future plans of the young individuals[3] when aged 16, close to the end of compulsory

[3]An unbiased representation of the educational aspirations of the BCS sample was particularly complicated to achieve. Out of the 11622 cohort members who took part in the age 16 follow-up only 6417 individuals responded to the school questions out of which just 4046 individuals could be grouped as having aspirations for academic, vocational or no post compulsory qualifications. The variable generated using this information could not be used in the statistical analysis because first, it could not be confidently considered representative of the total sample and second, it was decreasing the analytical sample to less than 4000 individuals. For that reason, the aspirations of the individuals who did not respond to the school questions, have been proxied, when available, from the parents' questionnaires at the age 10 and age 16

schooling, have been measured based on information on plans to continue for academic or vocational education or to leave education. The set up of the aspirations variable was based on questions asked of the individuals concerning their future plans. The individuals who were identified as aspiring to no post-compulsory education were the ones who mentioned that they are planning to leave education and did not mention any academic or vocational qualifications. As follows, the individuals who were identified for vocational qualifications were the ones who mentioned vocational qualifications but not an academic qualification. Finally, the individuals who were identified for academic qualifications included all individuals who mentioned academic qualifications as well as plans to go to the university and therefore superseding individuals who had also mentioned vocational qualifications. The questions were not asked in exactly the same order in the two cohorts; however the questions were very similar. Appendix A.1 represents the exact derivation procedure of the aspirations variables for the two cohorts using tree diagrams as well as details on some structural allocations that have been undertaken in the BCS aspirations variable.

Educational Choices

The final educational choices of the individuals have been measured after compulsory schooling, at the age of 26[4] for the BCS sample and at the age of 25[5] for the LSYPE sample. Educational choices, like educational aspirations, have been defined as the selection

follow-ups, asking the parents about the post-compulsory educational plans of the individual. Proxying the individuals aspirations with the parents aspirations allowed me to identify the aspirations of additionally 5,642 individuals. The questions answered by the parents concerned the individual's plans at the end of the term. The exact questions asked to the parents as well as the sources of these questions are included in Appendix A.1. To further validate this method an analysis of the similarity of parents' and individuals' aspirations for those individuals that both could be observed has been carried out. The outcomes from this analysis can be found in Appendix A.2. Also, a further robustness check in the regression analysis including only the individuals' response has been undertaken. The results and discussion of these estimations can be found in Appendix A.8. The analysis of the similarity between parents' and children's aspirations which show correspondence among parental and individual aspirations for over 50% of the observed individuals (as can be noticed from the percentages along the diagonal) as well as the correspondence of the robustness check in the regression analyses convinced me that parents' aspirations is a good proxy for individual's aspirations.

[4]The BCS age 26 follow-up had been carried out using mailed questionnaires and achieved a particularly low response (9003 cohort members). For that reason the missing observations of the age 26 follow-up have been completed using data from the age 29 follow-up when available. The educational choices variable has been derived from qualifications achieved when aged 26 using the coding frame provided by the UK Data Archive (BCS70 Twenty six-year Follow-up User Guide, Appendix 3: Coding frames for open-ended questions, pp.37-39). The additional data from the age 29 follow-up have been collected from the derived variable which was available in the dataset. The derivation of the age 26 highest qualification followed the same derivation procedure as the age 29 highest qualification which was available from the CLS (Dodgeon and Parsons, 2011).

[5]The post-compulsory educational choices of LSYPE's sample have been measured using data from Waves 4-8, capturing the highest post-compulsory qualification the individuals achieved from the year after compulsory schooling, at age 16-17, up until the age of 25. The derivation procedure of the educational choices variable followed the same derivation procedure as the BCS educational choices variable.

between academic, vocational and no post-compulsory education. The educational choices of individuals, as opposed to educational aspirations, define the actual participation in each educational alternative and have been measured based on already achieved qualifications and university attendance. Appendix A.3 represents the exact derivation procedure of the educational choices variables for the two cohorts clarifying in which type of education each qualification has been included. Further, Appendix A.4 includes a full list of the qualifications that have been considered for each of the two cohorts verifying the similarity and comparability of the two educational choices variables.

Key Variables

Individual ability

Both datasets include information on individuals' cognitive skills developed at early ages. The ability indicator of the BCS sample has been obtained at the age of 10 from the Friendly Maths test which has been developed especially for use in the survey. The test has been designed by the University of Bristol in collaboration with specialists in primary mathematics and has been completed by 11633 children. As mentioned above, although the LSYPE itself does not contain any information on the cohort members' cognitive skills, the dataset was matched with the NPD where the pupils' KS2 Maths scores taken at the age of 10-11 are available and have been used as indicators of their cognitive skills. The KS2 Maths scores were available for 14173 individuals. The ability measures of both cohorts precede entry into secondary school and the individuals' eventual educational achievement level.

The Friendly Maths test consisted of a full range of mathematical competence including awareness of number operations, arithmetic, number skills, fractions, algebra, geometry and statistics (Parsons, 2014). The KS2 Maths Test consisted of three separate tests covering both arithmetic and mathematical reasoning including fractions, algebra, measurement skills, geometry, ratio and proportion and statistics all within the cognitive domain (Standards and Testing Agency, 2012). Taking into account the content, purpose and the age of pupils taking the tests, LSYPE's KS2 Maths test have been considered strongly comparable with the BCS Friendly Maths test. Because the ability tests were not exactly identical it was not possible to use a raw test score for the analysis and for that reason the test scores have been rescaled to have a mean of zero and a standard deviation (sd)

of one which allowed performance to be reported on a consistent scale for pupils from both cohorts[6]. The distribution of each of the two tests for the analytical sample of each cohort is shown below in Figures 2.1 and 2.2 for the BCS and LSYPE respectively. From the distribution of the two tests it is revealed that although the results on the tests are not identical, they are similarly distributed for the BCS and LSYPE suggesting a fair comparability between the two ability components.

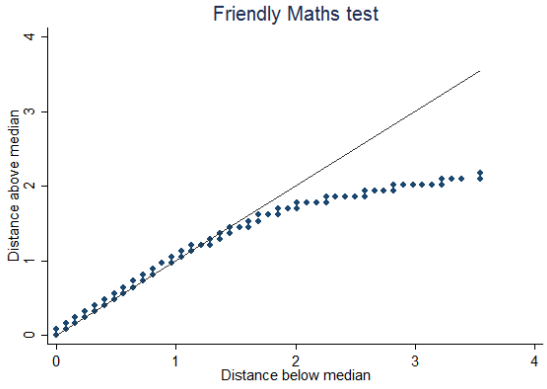


Figure 2.1: Distribution of test (BCS)

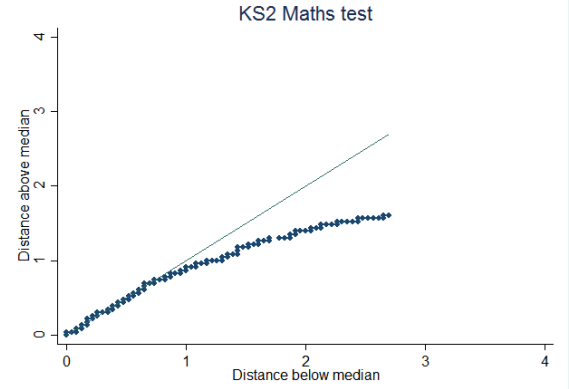


Figure 2.2: Distribution of test (LSYPE)

Socio-economic background

Parental occupation

Both datasets contain information on both parents' social class as estimated by their occupational category. The BCS social class estimation is based on the Registar General's Social Class (RGSC) with data collected from the age 10 follow up. The LSYPE social class estimation is based on the National Statistics Socio-Economic Classification (NSSEC) with data from Wave 1 (age 13-14). The occupational categories have been classified identically in the two samples and include professional, managerial and technical, skilled manual, skilled non-manual, partly skilled and unskilled occupations as well as unemployment. An additional category capturing missing information on parents' occupation is also included for both parents and for both cohorts.

Parental education

The measurement of parents' education was based on both parents' highest educational level based on qualifications achieved. The educational categories include academic degrees,

[6]For the standardisation of test scores the total sample for whom information was available has been used and not only the individuals in the analytical sample. Therefore, the standardised test scores reflect the ability position of individuals in the analytical sample compared to the total sample.

academic qualifications, other qualifications[7] and no qualifications. As with the occupation variable, an additional category capturing missing information on parents' education is also included for both parents and for both cohorts.

Control Variables

In addition to the key variables described above, both datasets provide rich information for the individuals and their households that allowed me to control for a large number of exogenous factors as required to identify only a ceteris paribus link from socio-economic background and cognitive skills to educational aspirations and choices[8]. The control variables combine individual demographics including gender and ethnicity and family composition characteristics including parents' age and number of children in the household. The set of control variables also includes the geographic area (urban/ rural indicator). A full list and a detailed description of the sources and composition of each of the control variables is included in Appendix A.5 for both cohorts. Some minor structural allocations that have been undertaken to some of the variables are also described.

2.3.4 Descriptive Statistics

The descriptive statistics of the analytical sample for both cohorts can be found in Appendix A.6. The full set of descriptive statistics comparing the total and the analytical samples of both cohorts are included in Appendix A.7. As mentioned, under some structural allocations to some of the variables the analytical sample remained highly representative of the total sample.

The descriptive statistics indicate that the LSYPE sample is aspiring more and has more education than the BCS sample, as expected. Furthermore, the LSYPE sample has more educated parents than the BCS sample, especially with regards to maternal education, and also the social class structure has changed somewhat between the two cohorts with

[7]Other Qualifications include: BCS: O levels, apprenticeship, SRN, other. LSYPE: GCSE, Level 1 qualifications, other.

[8]It is acknowledged that the school the individuals were attending is likely to play an important role in the formation of aspirations and the educational decisions of students. Introducing school fixed effects would limit the bias caused from not taking into account this important confounding factor. Unfortunately, the BCS dataset does not provide information regarding the school the individuals were attending. Nevertheless, the identification of the causal effect of socio-economic background and ability on aspirations and choices is beyond the scope of this chapter. The chapter aims to identify the changing influence of these effects from the older to the more recent cohort and I am able to do this, while minimising the upward bias in the estimation through the inclusion of important background characteristics. Within school peer effects on educational aspirations will be exclusively analysed in the next chapter.

a significant decrease in the proportion of unemployed mothers and an increase in the proportion of parents from higher social class occupations. This is particularly true for mothers in managerial occupations who increased from 12% in BCS to 24% in LSYPE. In terms of other background characteristics, there seems to be a greater proportion of ethnic minority students in LSYPE than in BCS and also parental age at individual's birth is increased by 2 to 3 years on average[9].

2.3.5 Educational aspirations and choices

The purpose of this section is to explain how the educational aspirations of each cohort near to the end of compulsory schooling are linked with their actual educational choices. The importance of looking both at educational aspirations and choices is that it allows a comprehensive investigation as to where the socio-economic effects emerge and to what extent these effects impact upon the actual choices of ambitious and non-ambitious individuals. The educational aspirations and choices of the BCS and the LSYPE cohorts are presented in Figures 2.3 and 2.4 respectively.

It is evident from the data that there is a considerable shift between both educational aspirations and choices of the older cohort compared to these of the most recent cohort. The most interesting and significant change between the two is on the increasing aspirations and actual participation in academic education. While in the BCS cohort 34% of the sample aspired to follow academic education and just 31% actually ended up having an academic qualification, in the LSYPE cohort the individuals were not only found to be much more ambitious for their future education with 74% of them aspiring academic education but also there is a significant increase in the actual participation as well. Overall, 67% of the individuals ended up having an academic qualification by the age of 25 which suggests a considerable expansion of academic education and a decline of any possible influence that could be limiting attendance in the past.

Further, the data reflect a decline of vocational education in terms of aspirations and a collapse in terms of actual participation. For 49% of the individuals in the BCS cohort, a vocational qualification was their highest achieved qualification by the age of 26, while for the LSYPE cohort this value decreased to just 25%. What is of particular interest is that

[9]The bigger difference that can be observed in average parental age between the two cohorts is because of the time they are observed. Parental age for the BCS sample is estimated at cohort members' birth while for the LSYPE sample parental age is measured at Wave 1 of the survey, when the young people were aged 13-14.

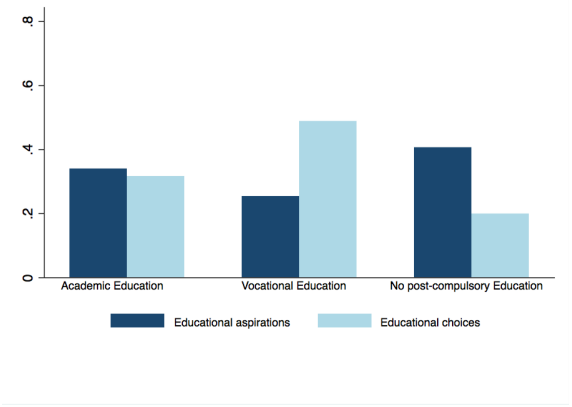
the decreasing participation in vocational education from the BCS to the LSYPE cohort appears not be arising from falling aspirations for vocational education but from increasing aspirations and actual participation in academic education. This is reflected from the fact that although a significant participation in vocational education can be identified in the BCS cohort, when looking at the individuals' aspirations, just 25% of them aspired to follow vocational education, implying that vocational education was not the direct goal of the individuals from the beginning[10]. Considering the LSYPE sample, only 23% of the individuals aspired to follow vocational education and, in contrast with the BCS cohort which finally decided to study for a vocational qualification, no more than 25% of the individuals ended up there.

What can be also identified is that not only the individuals who did not aspire to any post-compulsory education decreased from 41% to less than 4% but also the individuals who remained with no post-compulsory qualifications are significantly reduced from 20% to just 8%. These values reflect a significant expansion of both ambition for more education and of actual attainment, which, however, is not evenly spread across the various options that England's educational system provides and the labour market needs.

2.3.6 Whose aspirations for academic education are fulfilled?

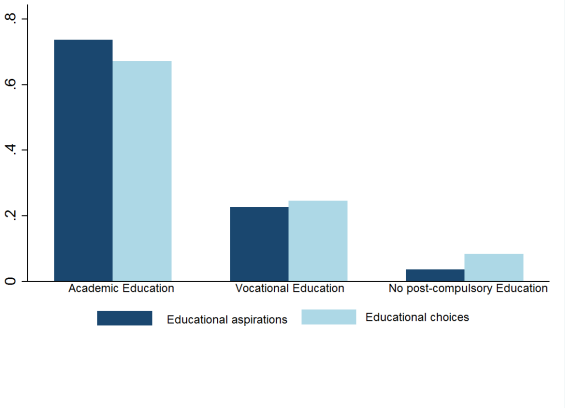
Following the analysis on aspirations and actual participation, this section discusses the social class and ability level of the individuals who aspired to academic education and managed to fulfil their aspirations by achieving an academic qualification. Each cohort has been divided in three ability tertiles (based on their achievement on the Friendly Math's test for the BCS cohort and their achievement on the KS2 Maths test for the LSYPE cohort.) defined as lower, middle and upper level and their social class is defined by their

[10]The cell of individuals in the older sample who aspired to leave education but ended up having a vocational qualification is the one where aspirations mostly differed from eventual choices. For that reason we were concerned that this might be an issue arising from the fact that the missing aspirations of the BCS cohort were proxied from their parents aspirations. We examined further this issue to see whether the aspirations of individuals in this particular cell have been mostly proxied from their parents and this is why their aspirations differ that much from their choices. In the analytical sample there are 1521 observations who aspired to leave education but ended up having a vocational qualification. From these, 471 responses (31%) came from the individuals and 1050 responses (69%) came from the parents. The aspirations variable has 6482 observations which 47% come from the individuals and 53% come from the parents, adjusting this weight, then, from these 1521 observations 35% come from the individual and 65% come from the parents. Considering these values, indeed this particular cell of individuals (aspiring to leave but doing vocational) is mostly taken from the parents' questionnaire rather than the individual's. Although that could be a reason for wrongly attributing the individual as having no aspirations for vocational education, still, the gap between vocational aspirations and choices is quite large and therefore enough to convince that young individuals were not aspiring to follow vocational education.



Academic Qualifications: Aspiration: 34% Choice: 31%,
Vocational Qualifications: Aspiration: 25% Choice: 49%
No post-compulsory Qualifications: Aspiration: 49% Choice: 40%
N=6652

Figure 2.3: Aspirations and choices (BCS)



Academic Qualifications: Aspiration:74% Choice: 67%,
Vocational Qualifications: Aspiration: 23% Choice: 25%
No post-compulsory Qualifications: Aspiration: 4% Choice: 8%
N=9405

Figure 2.4: Aspirations and choices (LYSPE)

parents' occupation and education. Figures 2.5 and 2.6 and Figures 2.7 and 2.8 show how the proportion of fulfilled academic aspirations of the BCS cohort varies as the individuals' social class and ability level vary estimating social class using parents' occupation and education respectively. Similarly, Figures 2.9 and 2.10 and Figures 2.11 and 2.12 show the same information for the LSYPE cohort.

From the BCS cohort, 68% of the individuals who aspired to academic education managed to fulfill this aspiration while from LSYPE the individuals who managed to fulfill their aspiration increased to 84%. The diagrams below show that the proportion of individuals with fulfilled academic aspirations is increasing as their ability level increases but in the LSYPE cohort there is less variation between the different levels suggesting that ability became less important for fulfilling the individuals' aspirations. On the other hand, although in both cohorts the proportion of individuals who have fulfilled academic aspirations increases as the social class of their parents goes up it can be seen that in the LSYPE cohort the proportion of fulfilled aspirations becomes much steeper across the occupational and educational categories suggesting that there are decreasing socio-economic effects. In addition, the lower proportion of individuals with fulfilled aspirations in the BCS cohort is 0.2 while in the LSYPE cohort this value is increased to 0.7 suggesting that there is a greater extent of fulfilled academic aspirations.

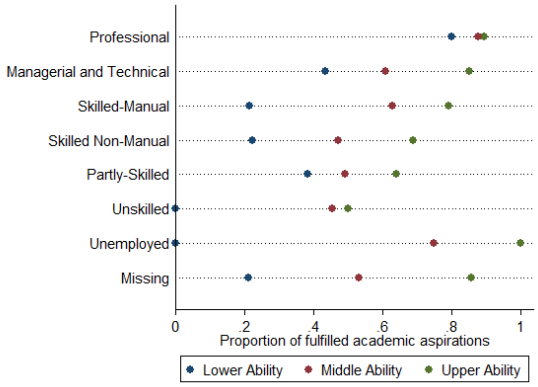


Figure 2.5: Father's occupation (BCS)

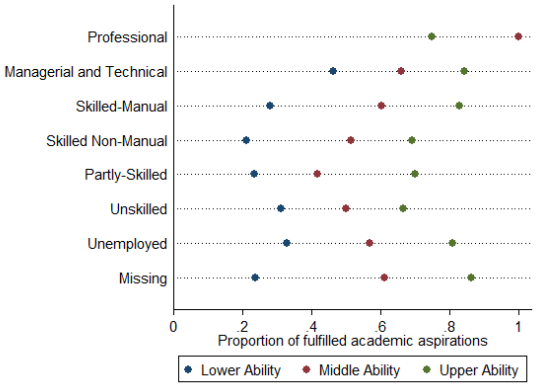


Figure 2.6: Mother's occupation (BCS)

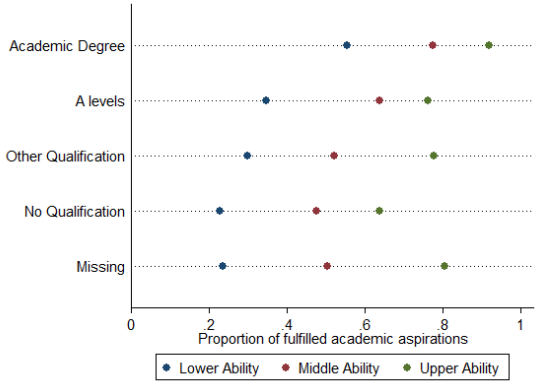


Figure 2.7: Father's education (BCS)

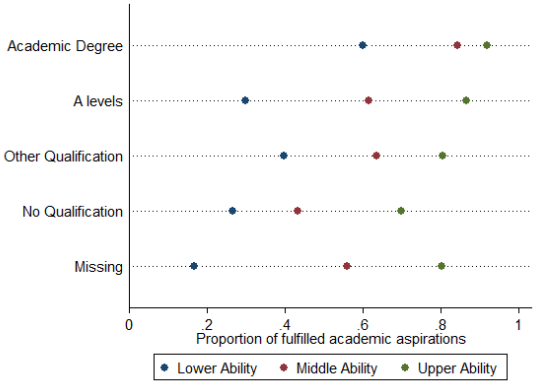


Figure 2.8: Mother's education (BCS)

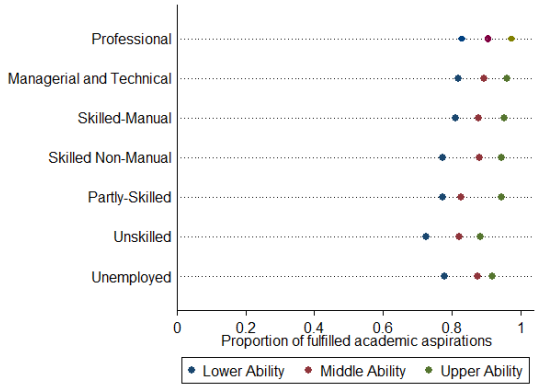


Figure 2.9: Father's occupation (LSYPE)

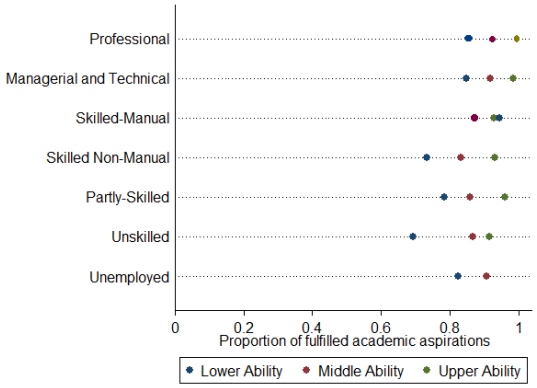


Figure 2.10: Mother's occupation (LSYPE)

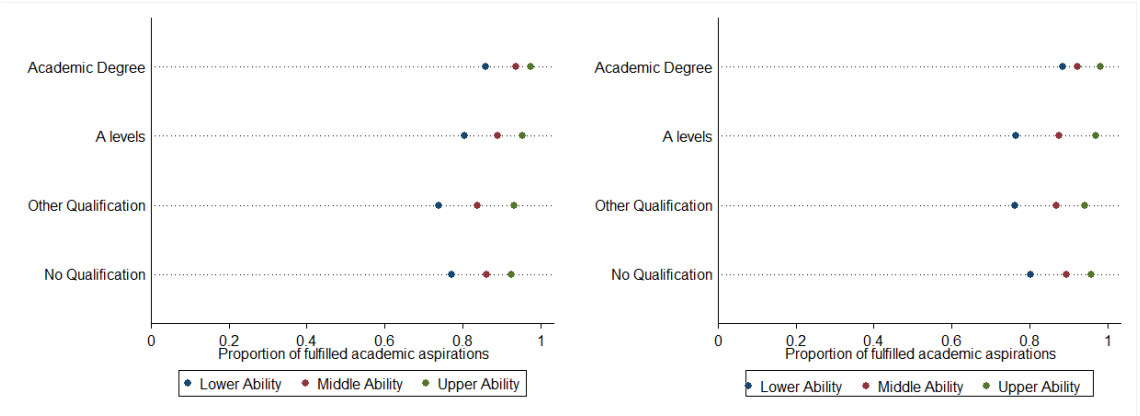


Figure 2.11: Father's education (LSYPE) **Figure 2.12: Mother's education (LSYPE)**

2.4 Methodology

2.4.1 Introduction to Unordered Multiple Choice Models

The purpose of this chapter is to investigate the changing influence of socio-economic background and ability on the educational aspirations and choices of young individuals. Empirical models for educational aspirations and choices are estimated using two different datasets and controlling separately for parents' education and parents' occupation in order to avoid multicollinearity among the independent variables.

The dependent variables which estimated educational aspirations and choices are unordered, having no natural ranking across the alternatives. Unordered choice models are motivated by a random utility model. All alternative choices are labelled arbitrarily and each individual chooses only one of the possible options. For each individual i and possible alternative k there is an unobserved random variable defined as a continuous latent variable $y^*_{i,k}$. This latent variable $y*_{i,k}$ conditional on a set of independent and control variables, x, is distributed for the i^{th} individual who has to choose between $j = 1, 2...k$ choices. Utility, conditional on the set of the independent and control variables is specified as:

$$y^*_{i,k} = \beta_k{}' x_i + \epsilon_{i,k}$$

The empirical model is driven by the probability that choice j is made meaning that if individual i makes choice j then one assumes that $y*_{i,k}$ is the maximum utility among the j options.

2.4.2 The Multinomial Logit Model

The technique that has been used for the three outcome unordered models is the Multinomial Logistic regression using Maximum Likelihood (ML) estimation techniques to estimate the parameters that best fit the data. The dependent variable is a categorical variable with individual i's chosen educational alternative k and although the independent variables do not vary across alternatives, the parameter β_j differs across them (Schmidheiny, 2007). In the method of ML, the parameter values which maximise the likelihood, or equivalently the log-likelihood, are picked and estimated using the Newton-Raphson iterative method

(Czepiel, 2012).

In Multinomial Logit Models (MLM), choice is a function of the characteristics of the individual making the choice and the explanatory variables remain constant over the alternative choices. For the specific MLM and particularly for the interpretation of the marginal effects which will be explained in detail below, for each educational alternative k the non-educational alternative k has been chosen as the reference category. As a result, the estimation procedure for aspirations allowed us to model the factors that affect the probability of aspiring academic education rather than not aspiring academic education, aspiring vocational education rather than not aspiring vocational education and not aspiring any post-compulsory education rather than aspiring post-compulsory education. Similarly, the estimation procedure for choices allowed us to model the factors that affect the probability of choosing academic education rather than not choosing academic education, choosing vocational education rather than not choosing vocational education and not choosing any post-compulsory education rather than choosing post-compulsory education.

The MLM analyses individual choice among discrete alternatives with the assumption that each individual i chooses the alternative that yields higher utility or satisfaction. For this specific estimation the following data model is estimated for the i^{th} individual for educational choice k:

$$Y^*_{i,k} = \beta_0 + \beta_{1,k} SES_i + \beta_{2,k} Ability_i + \gamma_k `X'_i + \epsilon_{i,k}$$

where the variables are:

- $Y^*_{i,k}$: the latent variable corresponding to educational aspiration or choice k of individual i
- β_0: the intercept parameter (constant).
- SES_i: the socio-economic component of individual i including occupation of both parents or highest educational achievement of both parents.
- $Ability_i$: the ability component of individual i.
- X_i: a vector of several controls for individual i including gender, ethnicity, parents' age, number of children in the household and whether living in an urban area. The controls include binary, categorical and continuous variables.
- $\epsilon_{i,k}$: the error component

for k indicating academic, vocational and no post-compulsory qualifications.

The latent variable $Y^*_{i,k}$ can be thought of as the utility associated with individual i choosing educational alternative k where there is some randomness in the actual amount of utility obtained which accounts for other unobserved factors that go into the choice. The value of the actual variable Y_i is then determined non-randomly from these latent variables as the randomness has been moved from the observed outcomes into the latent variables. Educational outcome k is then chosen if only the associated utility which is determined by the value of $Y_{i,k}$ is found to be greater than the utilities of all the other alternatives.

That is:

$$Pr(Y_i = 1) = Pr(max(Y_{i,1}, Y_{i,2}, Y_{i,3}) = Y_{i,1})$$

$$Pr(Y_i = 2) = Pr(max(Y_{i,1}, Y_{i,2}, Y_{i,3}) = Y_{i,2})$$

$$Pr(Y_i = 3) = Pr(max(Y_{i,1}, Y_{i,2}, Y_{i,3}) = Y_{i,3})$$

The dependent variable distinguishes how the likelihood of the educational aspiration or choice of an individual varies as the independent variables vary. The error component, $\epsilon_{i,k}$, represents any other unobserved factors that have an effect on educational choices. Table 2.3 below summarises the dependent and key variables used in each estimated model. As mentioned above, the estimated regression models include a number of other control variables apart from the socio-economic background and ability component. The same control variables among the two datasets have been used in order to make their results comparable. The ability component differs among the two datasets but in both cases represents the level of cognitive ability developed in early ages.

Table 2.3: Summary of estimated models

Model	Age	Dependent Variable	SES component	Ability component
BCS				
1	16	Educational aspirations	Parents' occupation	Friendly Maths Test (age 10)
2	16	Educational aspirations	Parents' education	Friendly Maths Test (age 10)
3	26	Educational choices	Parents' occupation	Friendly Maths Test (age 10)
4	26	Educational choices	Parents' education	Friendly Maths Test (age 10)
LSYPE				
5	16	Educational aspirations	Parents' occupation	KS2 Maths (age 10-11)
6	16	Educational aspirations	Parents' education	KS2 Maths (age 10-11)
7	25	Educational choices	Parents' occupation	KS2 Maths (age 10-11)
8	25	Educational choices	Parents' education	KS2 Maths (age 10-11)

2.4.3 Marginal Effects

In a MLM the sign and value of an estimated coefficient determines a log-odds ratio and when in that form is not as clear in determining the relationship between an independent variable and a dependent variable. For clear interpretations about the direction and magnitude of the relationship between an independent and a dependent variable in a MLM, marginal effects should be calculated and their standard errors (Bowen and Wiersema, 2004). The marginal effects are defined as the slope of the prediction function at a given value of the explanatory variables and thus inform us about the change in predicted probabilities due to a change in a particular predictor (Wulff, 2015). There are two different approaches of measuring marginal effects. The first is to set all of the predictors to their mean values resulting in marginal effects at the mean (MEM). The disadvantage of using this approach is that it is unlikely that there is a unit in the sample that is average on all model variables. In order to avoid this, the marginal effects have been estimated using average marginal effects (AME) which relies on actual values of the independent variables. The marginal effect is calculated for each individual according to their characteristics, and then averaged across all individuals.

The estimated marginal effects are surrounded by 95% confidence intervals. As referred above, the marginal effect shows the outcome of a unit change in each variable on the probability of choosing each educational alternative and in the specific case it is not interpreted relative to a reference category. In other words, the marginal effect for each of the regressors is examined on the probability of observing each of the three alternative outcomes, including the choice between academic, vocational and no post-compulsory qualifications. All categorical variables fitting in the model have been treated as factor variables and the marginal effect has been computed as a discrete change in the probability of having each characteristic rather than having the omitted category characteristic.

2.4.4 Selection in educational alternatives and omitted variable bias

Making use of the longitudinal nature of both datasets the chapter includes rich control variables that allowed to take into account a large number of exogenous factors as required to identify only a ceteris paribus link from socio-economic background and cognitive skills to educational aspirations and choices. These specific control variables could have a direct or an indirect impact on educational aspirations and choices and have been selected to be

used in the estimation as they can be considered exogenous to the individuals' aspirations and choices but are still likely to be highly associated with educational decisions. The set of explanatory variables includes family characteristics (number of children in the household and parents' age), individuals' demographic characteristics (gender, ethnicity) and a description of the area of residence (urban or rural). These variables remain the same in all models and for both surveys. Positive values in the marginal effect of each variable indicate that the probability of attending each type of education (or not attending any education) increases when an individual has that specific characteristic whereas negative values indicate that attendance to that type of education is reduced with that covariate.

It is widely acknowledged that the educational aspirations and choices of students can be influenced by a myriad of factors. For example the role of peers in influencing educational aspirations and choices, which is extensively examined in the next chapter, or the role of the school, which is discussed in more detail in Chapter 4, are not considered in this analysis. Further, selection in each educational alternative is one of the main econometric issues associated with causal estimations. It is possible that part of the estimated socio-economic and ability effect could be capturing other unobserved characteristics of the individual, different from those that are included in the set of control variables. As identified in previous analyses following a similar approach, by "simply including additional observed variables cannot definitely eliminate omitted variable bias arising from unobservable factors and in the absence of a randomized experiment, there is a limit to how far this study can go in establishing causal relationships" (Vignoles et al., 2011, p. 5).

To identify the causal effect of socio-economic background and ability on educational aspirations and choices is beyond the scope of this chapter. The chapter follows a strongly comparative approach to examine, *between* cohorts, the change in the importance of socio-economic background and ability in influencing educational attainment and *within* cohorts, the difference in the effect of these inputs between aspirations and choices. To the extent that biases are the same across the compared models, they will cancel out when looking at these differences.

2.5 Empirical Results

This section presents and discusses the findings of the main empirical models that have been estimated using Multinomial Logistic regression. Table 2.4 describes the combination

of models as presented in the main regression tables, sorted by outcome of interest and socio-economic component.

Table 2.4: Model combinations

Case	Models	Age	Outcome	SES component
A	1 and 5	16	Aspirations	Parents' occupation
B	2 and 6	16	Aspirations	Parents' education
C	3 and 7	25-26	Choices	Parents' occupation
D	4 and 8	25-26	Choices	Parents' education

For the interpretation of the results average marginal effects have been computed and the main results showing the outcome of these estimations are reported in Tables 2.5, 2.6, 2.7 and 2.8. The results are divided into two Sections. Section 2.5.1 includes the estimates for educational aspirations at age 16 and Section 2.5.2 includes the estimates of educational choices at age 25 (LSYPE) and age 26 (BCS). Section 2.5.3 presents a summary of the main results. Post estimations on the main empirical models have been carried out, including specification tests and measurements of goodness of fit, the results of which are presented in Section 2.5.4.

2.5.1 Educational Aspirations

Case A: Educational aspirations measuring socio-economic background with parental occupation

Case A highlights the socio-economic and ability effects associated with educational aspirations of young individuals, as expressed close to the end of compulsory schooling, measuring socio-economic effects using parents' occupation. The marginal effects are presented in Table 2.5. As it is evident from these results for both samples, higher scores on the arithmetic test, implying higher levels of cognitive skills, are associated with higher aspirations to follow academic education and lower aspirations to follow vocational or no post-compulsory education. The results suggest that an individual's ability level had a lesser impact on educational aspirations in the more recent sample. A 1 sd increase on the arithmetic test scores in the older sample is associated with 18 percentage points (pp) higher probability of aspiring to attend academic education and 13 pp less probability not to attend any post compulsory education while in the the more recent sample, the ability effect decreased to 14 pp higher probability of aspiring to academic qualifications and to just 2 pp less probability not to aspire any post-compulsory education. In the case of

vocational education, the results imply a collapse of aspirations among the highest ability individuals, even when compared with no post-compulsory education. While in the older sample, a 1 sd increase on the arithmetic test scores was found to be associated with 5 pp less probability of aspiring vocational education, in the more recent sample this gap increased to 12 pp suggesting a significant downfall for vocational education and that it became the aspiration of the least able individuals.

Further, the results demonstrate that the socio-economic effect exists in both cohorts but in the more recent sample, parental occupation became less important in determining individuals' aspirations for post-compulsory education. For example, individuals with fathers in professional or managerial and technical occupations in the older sample are found to be 34 pp and 22 pp respectively more likely to aspire to academic education while in the more recent sample the effect is significantly decreased to 14 pp and 12 pp respectively. Further, it is observed that although paternal occupation has a significant impact on educational aspirations for both cohorts, maternal occupation appears to be mostly insignificant for their children's aspirations.

The most interesting and striking finding is on the change of the socio-economic effect, as observed mainly from paternal occupation, on vocational aspirations from the older to the more recent cohort. While in the older sample individuals with fathers in higher social class occupations, especially those in professional, managerial and technical and skilled occupations, appear to be more likely to aspire to follow vocational education, individuals in the more recent cohort with parents in higher social-class occupations appear to be less likely to aspire vocational education. This finding suggests a collapse of aspirations for vocational learning among the highest socio-economic families and among the more able individuals.

Case B: Educational aspirations measuring socio-economic background with parental education

Table 2.6 presents the same information, measuring socio-economic background using parental education. The results follow a similar pattern as when estimating socio-economic background using parents' occupation and verify the robustness of these findings. There is a significant decline in the importance of cognitive ability in influencing educational aspirations for academic and no post-compulsory education, while this effect becomes more negative for vocational aspirations.

The results indicate that parental education has a significant effect on their children's educational aspirations. It can be identified that, in both cohorts, individuals with more educated parents are more likely to aspire to academic education and significantly less likely to aspire to vocational and no post-compulsory education. Interestingly, although both paternal and maternal education seem to influence their children's aspirations, maternal education appears to have a stronger impact as opposed to the findings in Case A where paternal inputs seemed to be more important.

Nevertheless, the magnitude of the socio-economic effect is significantly reduced from the older to the more recent sample. In the older cohort, individuals with mothers holding an academic degree were 25 pp more likely to aspire to follow academic education compared to those whose mother had no qualifications, while in the more recent sample this effect is significantly reduced to 12 pp. Similarly, students in the older sample whose mother studied for A levels were 19 pp more likely to aspire to the academic route while this value decreased to just 6 pp in the more recent sample. The probability of not aspiring to stay in education after completing compulsory participation at age 16 is significantly higher for individuals with less educated parents for both cohorts but the decline of the socio-economic effect is still apparent.

In the case of of vocational education, the socio-economic effect becomes more negative from the older to the more recent sample. In the BCS cohort individuals with more educated mothers were significantly more likely to aspire to follow vocational education while in the more recent sample the effect becomes negative. These findings support what has been discussed above about the declining aspirations to participate in vocational education as it appears that more well-off families in England are increasingly pushing their children away from vocational training.

All in all, the results show a significant decline of the socio-economic and ability effect on educational aspirations for academic and no post-compulsory education. Nonetheless, both the ability and the socio-economic effect increased from the older to the more recent sample for vocational education aspirations suggesting a dramatic downfall for vocational education's recognition.

2.5.2 Educational Choices

Case C: Educational choices measuring socio-economic background with parental occupation

While the previous section examines the changing importance of socio-economic background and ability in influencing educational aspirations, this section investigates the importance of these inputs on final choices based on qualifications that have been achieved. The estimations of Case C, measuring socio-economic effects using parents' occupation, are reported in Table 2.7.

The results continue to show the significant effect of cognitive ability in influencing final educational choices. The magnitude of the effect does not have a notable difference between aspirations and choices but it is still significantly decreased from the older to the more recent sample. Individuals in the BCS sample who scored a 1 sd higher on the arithmetic test were found to be 17 pp more likely to have studied for an academic qualification, 10 pp less likely to have studied for a vocational qualification and 7 pp less likely not to have followed any post-compulsory education. In the LSYPE sample the ability effect is decreased to 16 pp for academic qualifications and to just 4 pp for no post-compulsory qualifications while for vocational education the effect becomes more negative to 12 pp.

The influence of socio-economic background on academic and no post-compulsory education choices follows a very similar pattern as the one observed for educational aspirations, although it seems that the effect has a a slightly higher impact on final choices rather than aspirations. Students from disadvantaged backgrounds were less likely to have achieved an academic qualification and more likely to leave education without achieving any post-compulsory qualification and this effect is still of greater magnitude when observing paternal rather than maternal education. Comparing between the two cohorts, the importance of these inputs is significantly reduced from the older to the more recent sample. For example, individuals in the BCS sample with fathers in professional, managerial and technical and skilled-manual occupations are found to be 36 pp, 24 pp and 19 pp respectively, more likely to have achieved an academic qualification and 13 pp, 11 pp, 9 pp respectively, less likely not to have attended any post-compulsory education than individuals with fathers from unskilled occupations. In the LSYPE sample the magnitude of these effects is decreased with individuals with fathers in professional, managerial and technical and skilled-manual being 20 pp, 10 pp and 10 pp respectively, more likely to have achieved

an academic qualification and 3 pp, 2 pp and 2 pp respectively, not to have attended any post-compulsory education compared to individuals with fathers in unskilled occupations. These estimations as stated above, reveal a significant decline of the socio-economic effect on educational choices when considering academic and no post-compulsory education. The lower impact of socio-economic background on attaining no post-compulsory qualifications is likely to be a function of the general increase in educational attainment between the cohorts such that very few people in the most recent cohort have no post-compulsory qualifications by age 25.

For vocational education, unlike what has been observed when estimating educational aspirations, the impact of being from a socially advantaged family seems to have a negative effect on vocational choices in the older sample. This is still the case in the more recent sample although a slight decline in the importance of the effect can be observed contradicting the findings on educational aspirations. If this finding remains robust in Case D, estimating socio-economic effects using parental education, then it might be the case that although aspirations for vocational education are socially graded to a greater extent in the more recent cohort, when finally reaching post-compulsory education, schools might have started to do a better job towards guiding young individuals to study for vocational qualifications.

Case D: Educational choices measuring socio-economic background with parental education

The results of the final case, Case D, measuring the effect of socio-economic background, as estimated from parental education, and of cognitive ability on final educational choices are presented in Table 2.8. The effect of ability on educational choices is significant and follows a decreasing direction from the older to the more recent sample for academic and no post-compulsory education. Individuals in the BCS sample who scored a 1 sd higher on the arithmetic test were found to be 16 pp more likely to have attended academic education and 7 pp less likely not have attended any post-compulsory education while in the LSYPE sample these values are reduced to 16 pp and 4 pp respectively. The ability effect on vocational education becomes more negative from 10 pp in the older cohort to 12 pp in the more recent cohort.

Individuals with more educated parents are estimated to be significantly more likely to have studied for an academic qualification, and less likely to have studied for vocational or no post-compulsory qualifications. As observed in Case B, this input has a greater

impact on educational choices when received from the child's mother rather than the father. Individuals' in the older sample with mothers holding an academic degree were 21 pp more likely to follow the academic route while in the more recent sample this effect decreased to 14 pp The findings confirm the robustness of the previous estimations and suggest that an individuals' social background became less important in restricting individuals post-compulsory educational choices. It is worth mentioning that since in the more recent cohort mothers with degrees are less rare, and therefore less selected, it might be partly the reason why the maternal degree effects are lower.

In the case of vocational education, the socio-economic effect follows a similar pattern to the one observed when estimating the importance of this effect using parental occupation, contradicting the findings on educational aspirations. In the older cohort individuals with mothers holding an academic degree were 15 pp less likely to have followed the vocational route compared to individuals whose mother had no qualifications while this effect decreased to 7 pp in the more recent sample. As discussed above, it might be the case that schools after all might have started to guide students towards vocational qualifications in a more efficient way. Of course, to shed light with regards to this assumption further analysis would be required.

Table 2.5: MLM Marginal Effects on educational aspirations measuring SES with parental occupation

	BCS(1986)			LSYPE(2006)		
	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx
Cognitive ability						
Arithmetic test	0.183***	-0.050***	-0.133***	0.138***	-0.117***	-0.021***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.002)
Socio-economic background						
Father Professional	0.337***	0.138***	-0.475***	0.138***	-0.131***	-0.006
	(0.047)	(0.047)	(0.048)	(0.035)	(0.036)	(0.019)
Father Managerial	0.219***	0.091**	-0.310***	0.117***	-0.113***	-0.004
	(0.043)	(0.040)	(0.036)	(0.027)	(0.026)	(0.013)
Father Skilled-Man	0.176***	0.079*	-0.254***	0.102***	-0.100***	-0.002
	(0.045)	(0.043)	(0.039)	(0.030)	(0.030)	(0.015)
Father Skilled Non-Manual	0.097**	0.070*	-0.167***	0.047*	-0.049*	0.003
	(0.043)	(0.039)	(0.035)	(0.026)	(0.025)	(0.012)
Father Partly Skilled	0.075	0.067	-0.142***	0.074**	-0.076***	0.001
	(0.046)	(0.041)	(0.037)	(0.030)	(0.029)	(0.014)
Father Unemployed	-0.064	0.106	-0.042	-0.004	-0.011	0.015
	(0.085)	(0.068)	(0.067)	(0.037)	(0.036)	(0.018)
Father Missing	0.140***	0.058	-0.199***	0.010	-0.022	0.012
	(0.045)	(0.042)	(0.038)	(0.026)	(0.025)	(0.012)
Mother Professional	0.151	0.139	-0.290	0.325	0.159	-0.483
	(0.109)	(0.144)	(0.205)	(9.820)	(21.274)	(11.454)
Mother Managerial	0.109***	0.017	-0.125***	0.068***	-0.043**	-0.025***
	(0.027)	(0.027)	(0.028)	(0.021)	(0.021)	(0.008)
Mother Skilled-Manual	0.038	0.009	-0.047**	0.025	0.004	-0.029***
	(0.025)	(0.024)	(0.024)	(0.021)	(0.020)	(0.008)
Mother Skilled Non-Manual	0.001	0.048*	-0.049*	-0.012	0.025	-0.013*
	(0.030)	(0.028)	(0.029)	(0.022)	(0.021)	(0.008)
Mother Partly Skilled	-0.016	0.007	0.009	-0.010	0.024	-0.014*
	(0.026)	(0.024)	(0.024)	(0.021)	(0.020)	(0.007)
Mother Unemployed	0.059**	-0.013	-0.046*	0.063***	-0.036	-0.027***
	(0.025)	(0.024)	(0.023)	(0.024)	(0.024)	(0.010)
Mother Missing	0.037	-0.012	-0.025	0.103***	-0.063**	-0.040***
	(0.029)	(0.029)	(0.029)	(0.032)	(0.031)	(0.015)
Control variables						
Girl	0.008	0.097***	-0.105***	0.115***	-0.101***	-0.014***
	(0.010)	(0.010)	(0.011)	(0.008)	(0.008)	(0.004)
White	-0.358***	0.012	0.346***	-0.284***	0.224***	0.060***
	(0.028)	(0.033)	(0.037)	(0.011)	(0.012)	(0.008)
Father's Age	0.000	0.001***	-0.001***	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Mother's Age	0.003***	-0.002*	-0.001	0.006***	-0.005***	-0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Children in household	-0.023***	-0.005	0.027***	-0.019***	0.016***	0.003*
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.001)
Urban	0.015	0.014	-0.030*	-0.008	0.005	0.003
	(0.015)	(0.016)	(0.017)	(0.010)	(0.010)	(0.005)
Observations	6652	6652	6652	9405	9405	9405

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Omitted groups: *Parental Occupation:* Unskilled *Gender:* Boy *Ethnicity:* Other *Area:* Rural

Table 2.6: MLM Marginal Effects on educational aspirations measuring SES with parental education

	BCS(1986)			LSYPE(2006)		
	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx
Cognitive ability						
Arithmetic test	0.175***	-0.049***	-0.126***	0.134***	-0.114***	-0.020***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.002)
Socio-economic background						
Father Degree	0.176***	0.005	-0.182***	0.094***	-0.072***	-0.021**
	(0.017)	(0.022)	(0.024)	(0.016)	(0.016)	(0.010)
Father A levels	0.089***	0.032	-0.121***	0.026*	-0.016	-0.010
	(0.018)	(0.020)	(0.022)	(0.015)	(0.015)	(0.008)
Father Other Qualification	0.032**	0.008	-0.041***	0.010	-0.010	-0.000
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.006)
Father Missing	0.026	-0.016	-0.011	-0.034***	0.026**	0.008
	(0.019)	(0.020)	(0.020)	(0.012)	(0.012)	(0.006)
Mother Degree	0.245***	0.094***	-0.339***	0.123***	-0.095***	-0.028***
	(0.027)	(0.036)	(0.047)	(0.014)	(0.014)	(0.007)
Mother A levels	0.191***	0.082***	-0.273***	0.057***	-0.027*	-0.029***
	(0.026)	(0.031)	(0.038)	(0.015)	(0.014)	(0.008)
Mother Other Qualification	0.080***	0.038***	-0.118***	0.025**	-0.012	-0.013***
	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.005)
Mother Missing	0.023	0.028	-0.051**	0.083***	-0.072***	-0.011
	(0.020)	(0.021)	(0.021)	(0.024)	(0.024)	(0.011)
Control variables						
Girl	0.008	0.097***	-0.106***	0.115***	-0.101***	-0.014***
	(0.010)	(0.010)	(0.011)	(0.008)	(0.008)	(0.004)
White	-0.349***	0.009	0.340***	-0.306***	0.239***	0.067***
	(0.027)	(0.033)	(0.036)	(0.011)	(0.011)	(0.008)
Father's Age	0.000	0.001***	-0.001***	0.001	-0.001	-0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Mother's Age	0.003***	-0.002	-0.001	0.005***	-0.004***	-0.001**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Children in household	-0.021***	-0.004	0.024***	-0.016***	0.014***	0.002
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.001)
Urban	0.011	0.009	-0.020	-0.002	0.002	0.000
	(0.015)	(0.016)	(0.016)	(0.010)	(0.010)	(0.005)
Observations	6652	6652	6652	9405	9405	9405

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Omitted groups: *Parental education:* No qualifications *Gender:* Boy *Ethnicity:* Other *Area:* Rural

Table 2.7: MLM Marginal Effects on educational choices measuring SES with parental occupation

	BCS(1986)			LSYPE(2006)		
	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx
Cognitive ability						
Arithmetic test	0.172***	-0.103***	-0.069***	0.159***	-0.119***	-0.040***
	(0.005)	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)
Socio-economic background						
Father Professional	0.361***	-0.234***	-0.127***	0.200***	-0.173***	-0.026
	(0.047)	(0.052)	(0.039)	(0.041)	(0.042)	(0.029)
Father Managerial	0.242***	-0.134***	-0.108***	0.097***	-0.074***	-0.023
	(0.044)	(0.042)	(0.028)	(0.029)	(0.028)	(0.019)
Father Skilled-Manual	0.194***	-0.109**	-0.085***	0.095***	-0.076**	-0.020
	(0.046)	(0.045)	(0.031)	(0.033)	(0.032)	(0.022)
Father Skilled Non-Manual	0.111**	-0.055	-0.056**	0.013	-0.011	-0.002
	(0.044)	(0.041)	(0.026)	(0.029)	(0.027)	(0.018)
Father Partly Skilled	0.088*	-0.063	-0.025	0.049	-0.053*	0.005
	(0.046)	(0.044)	(0.028)	(0.032)	(0.031)	(0.020)
Father Unemployed	0.085	-0.101	0.016	-0.040	0.015	0.026
	(0.073)	(0.073)	(0.047)	(0.039)	(0.037)	(0.024)
Father Missing	0.151***	-0.094**	-0.057**	-0.022	0.004	0.018
	(0.045)	(0.044)	(0.029)	(0.029)	(0.027)	(0.018)
Mother Professional	0.066	-0.138	0.072	0.179***	-0.091*	-0.087*
	(0.082)	(0.143)	(0.122)	(0.052)	(0.054)	(0.053)
Mother Managerial	0.092***	-0.092***	-0.000	0.088***	-0.060***	-0.028**
	(0.026)	(0.030)	(0.024)	(0.023)	(0.022)	(0.014)
Mother Skilled-Manual	0.049**	-0.050*	0.000	0.037	-0.020	-0.017
	(0.024)	(0.026)	(0.021)	(0.023)	(0.021)	(0.013)
Mother Skilled Non-Manual	-0.023	0.002	0.020	0.001	-0.019	0.018
	(0.030)	(0.032)	(0.024)	(0.024)	(0.023)	(0.014)
Mother Partly Skilled	-0.037	0.021	0.016	0.006	-0.009	0.004
	(0.025)	(0.027)	(0.020)	(0.023)	(0.022)	(0.013)
Mother Unemployed	0.039	-0.048*	0.008	0.050**	-0.041*	-0.009
	(0.024)	(0.026)	(0.020)	(0.026)	(0.024)	(0.015)
Mother Missing	0.025	-0.020	-0.005	0.130***	-0.094***	-0.036
	(0.028)	(0.032)	(0.025)	(0.034)	(0.033)	(0.022)
Control variables						
Girl	0.026***	-0.017	-0.009	0.118***	-0.081***	-0.037***
	(0.010)	(0.012)	(0.010)	(0.008)	(0.008)	(0.006)
White	-0.305***	0.212***	0.092***	-0.239***	0.158***	0.081***
	(0.027)	(0.036)	(0.029)	(0.011)	(0.011)	(0.008)
Father's Age	0.000	0.000	-0.001**	0.002**	-0.002*	-0.000
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Mother's Age	0.004***	-0.002*	-0.002**	0.007***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Children in household	-0.027***	0.000	0.027***	-0.026***	0.016***	0.010***
	(0.005)	(0.006)	(0.004)	(0.003)	(0.003)	(0.002)
Urban	-0.006	-0.000	0.006	-0.018	0.006	0.012
	(0.014)	(0.018)	(0.015)	(0.011)	(0.011)	(0.008)
Observations	6652	6652	6652	9405	9405	9405

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Omitted groups: *Parental occupation:* Unskilled *Gender:* Boy *Ethnicity:* Other *Area:* Rural

Table 2.8: MLM Marginal Effects on educational choices measuring SES with parental education

	BCS(1986)			LSYPE(2006)		
	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No post-comp. Education Mfx
Cognitive ability						
Arithmetic test	0.161***	-0.097***	-0.065***	0.155***	-0.118***	-0.037***
	(0.005)	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)
Socio-economic background						
Father Academic Degree	0.206***	-0.133***	-0.074***	0.122***	-0.070***	-0.052***
	(0.016)	(0.025)	(0.022)	(0.017)	(0.017)	(0.013)
Father A levels	0.112***	-0.023	-0.089***	0.024	-0.012	-0.012
	(0.018)	(0.024)	(0.022)	(0.016)	(0.016)	(0.011)
Father Other Qualification	0.047***	-0.021	-0.025**	0.032**	-0.006	-0.026***
	(0.013)	(0.015)	(0.012)	(0.014)	(0.014)	(0.009)
Father Missing	0.038**	-0.013	-0.025	-0.037***	0.028**	0.009
	(0.019)	(0.022)	(0.017)	(0.013)	(0.012)	(0.008)
Mother Academic Degree	0.205***	-0.148***	-0.057	0.136***	-0.072***	-0.065***
	(0.023)	(0.040)	(0.038)	(0.014)	(0.015)	(0.011)
Mother A levels	0.138***	-0.124***	-0.014	0.071***	-0.013	-0.058***
	(0.024)	(0.035)	(0.030)	(0.015)	(0.015)	(0.011)
Mother Other Qualification	0.090***	-0.050***	-0.040***	0.020*	0.008	-0.027***
	(0.011)	(0.014)	(0.012)	(0.012)	(0.011)	(0.007)
Mother Missing	0.022	0.002	-0.024	0.109***	-0.058**	-0.052***
	(0.020)	(0.024)	(0.019)	(0.025)	(0.026)	(0.018)
Control variables						
Girl	0.026***	-0.017	-0.009	0.119***	-0.082***	-0.037***
	(0.010)	(0.012)	(0.009)	(0.008)	(0.008)	(0.006)
White	-0.293***	0.210***	0.083***	-0.253***	0.160***	0.093***
	(0.026)	(0.036)	(0.029)	(0.011)	(0.011)	(0.008)
Father's Age	0.000	0.000	-0.001**	0.003***	-0.002**	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Mother's Age	0.004***	-0.002*	-0.002***	0.006***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Children in household	-0.026***	0.000	0.026***	-0.024***	0.015***	0.009***
	(0.005)	(0.006)	(0.004)	(0.003)	(0.003)	(0.002)
Urban	-0.010	-0.002	0.012	-0.011	0.004	0.007
	(0.014)	(0.018)	(0.015)	(0.011)	(0.011)	(0.008)
Observations	6652	6652	6652	9405	9405	9405

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Omitted groups: *Parental education:* No qualifications *Gender:* Boy *Ethnicity:* Other *Area:* Rural

2.5.3 Summary of Results

Table 2.9 and 2.10, summarise the main empirical results on the estimations of educational aspirations and choices of both cohorts when measuring socio-economic effects using parents' occupation and education respectively. The findings from the two socio-economic components show very similar results with regards to the direction of the effect from the older to the more recent sample verifying the robustness of these findings.

The results on aspirations and choices for academic and no-post compulsory education lead to very similar conclusions. It is indicated that the effect of ability on educational aspirations and choices for academic and no post-compulsory education is significantly decreased from the older to the more recent sample and this finding remains robust for both socio-economic components. Also, interestingly, the chapter identifies that the two socio-economic components have a different impact by parental gender. Paternal occupation seems to be a more important indicator of educational decisions compared to maternal occupation while on the other hand, maternal education seems to be more important than paternal education. This finding supports what is believed in the literature, that "the external effects on children associated with parental education are larger for maternal education than for paternal, because mothers tend to be the main provider of care within the household" (Chevalier et al., 2013, p. 2). All in all, these findings indicate that the expansion of academic education has proportionately benefited individuals from all social backgrounds.

In the case of vocational education, the findings are more complicated to interpret and for valid conclusions to be made. The data indicate that although aspirations for vocational education were much lower than those for academic education for both cohorts, actual participation was significantly higher in the older sample. The empirical findings indicate an increasingly negative socio-economic effect on vocational aspirations from the older to the more recent sample which, though, becomes slightly decreasing for final choices. Students from better-off families are found to be increasingly aspiring to not go into vocational education, though once they reach post-compulsory education, it seems that some of them ultimately choose to follow a vocational route to a greater extent in the more recent cohort. It could be that schools started doing a better job in the later cohort on persuading students that they would be better off in vocational education. On the other hand, the decreasing socio-economic effect identified on final choices might be driven

by specific vocational qualifications. In the future, further research will be carried out, ideally using datasets providing detailed information on vocational qualifications achieved, in order to enlighten on this matter.

Table 2.9: Summary of results using paternal occupation as the SES component

	BCS		LSYPE	
	Aspirations	**Choices**	**Aspirations**	**Choices**
Academic Education				
Arithmetic Test	0.183***	0.172***	0.138***	0.159***
	(0.005)	(0.005)	(0.004)	(0.004)
Father Professional	0.337***	0.361***	0.138***	0.200***
	(0.047)	(0.047)	(0.035)	(0.041)
Vocational Education				
Arithmetic Test	-0.050***	-0.103***	-0.117***	-0.119***
	(0.005)	(0.006)	(0.004)	(0.004)
Father Professional	0.138**	-0.234 ***	-0.123***	-0.173***
	(0.047)	(0.052)	(0.036)	(0.042)
No post-compulsory Education				
Arithmetic Test	-0.133***	-0.021***	-0.069***	-0.040***
	(0.005)	(0.005)	(0.002)	(0.003)
Father Professional	-0.447***	-0.127***	-0.006	-0.026
	(0.048)	(0.039)	(0.019)	(0.029)
Observations	6652	6652	9405	9405

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Omitted group: Occupation: Unskilled

Table 2.10: Summary of results using maternal education as the SES component

	BCS		LSYPE	
	Aspirations	**Choices**	**Aspirations**	**Choices**
Academic Education				
Arithmetic Test	0.175***	0.161***	0.134***	0.155***
	(0.005)	(0.005)	(0.004)	(0.004)
Mother Degree	0.245***	0.205***	0.123***	0.136***
	(0.027)	(0.023)	(0.014)	(0.014)
Vocational Education				
Arithmetic Test	-0.049***	-0.097***	-0.114***	-0.118***
	(0.005)	(0.006)	(0.004)	(0.004)
Mother Degree	0.094***	-0.148***	-0.095***	-0.072***
	(0.036)	(0.026)	(0.014)	(0.015)
No post-compulsory Education				
Arithmetic Test	-0.126***	-0.065***	-0.020***	-0.037***
	(0.005)	(0.005)	(0.002)	(0.003)
Mother Degree	-0.339***	-0.057***	-0.028***	-0.065***
	(0.047)	(0.038)	(0.007)	(0.011)
Observations	6652	6652	9405	9405

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Omitted group: Occupation: Unskilled

2.5.4 Specification Tests

Measurements of fit

A description of the overall test of the relationship between the dependent and independent variables is presented in Table 2.11. The model fitting information presented below reveals that for all the main regression models that have been estimated the p-values of the chi^2 statistics were 0.000, less than the level of significance of 0.05, suggesting that the relationship among the dependent and independent variables is statistically significant.

Table 2.11: Model fitting information

Model	Log-likelihood	Chi^2	df	Significance
BCS				
1	-7153.546	2813.819	44	0.000
2	-7057.009	3006.893	32	0.000
3	-6862.507	2312.716	44	0.000
4	-6813.435	2410.859	32	0.000
LSYPE				
5	-6662.649	2624.475	44	0.000
6	-5794.639	2815.930	32	0.000
7	-7500.390	12964.6767	44	0.000
8	-7486.924	2991.606	32	0.000

Strength of multinomial logistic regression

As the relationship among the dependent and independent variables within the main estimated models is found to be significant, the next step was to compute correlation measurements in order to estimate the strength of this relationship. When analysing data with Multinomial Logistic regression the model estimates ML estimations which are not calculated to minimise variance and as a result the OLS approach to goodness-of-fit cannot be applied. Although there is not an exact equivalent statistic to the R^2, in order to evaluate the goodness-of-fit on the MLM, pseudo-R^2 measures have been developed which will be applied on the estimated models. Pseudo-R^2 measurements are on a similar scale with R^2 measurements, ranging from 0 to 1, with higher values indicating better model fit (Scott Long, 1997).

In this case, following the example of Madhu et al. (2014), the Cox and Snell (or ML) pseudo-R^2 value and the Nagelkerke (or Cragg-Uhler) pseudo-R^2 value shown below provide a kind of indication of the amount of variation in the dependent variable and suggest the

amount of variability explained by the explanatory variables used in the estimated models. In reference to Table 2.12, Cox and Snell and Cragg-Uhler pseudo-R^2 values for the main estimated models using the BCS are found to be between 0.26 and 0.31 and between 0.30 and 0.36 respectively. As follows, Cox and Snell and Cragg-Uhler pseudo-R^2 values for LSYPE's regression models are found to be between 0.24 and 0.25 and between 0.30 and 0.32 respectively (rounded up to the nearest integer).

Table 2.12: Pseudo R^2

Model	Cox and Snell/ML	Cragg-Uhler/Nagelkerke
BCS		
1	0.298	0.337
2	0.315	0.356
3	0.259	0.296
4	0.268	0.307
LSYPE		
5	0.237	0.318
6	0.240	0.318
7	0.247	0.301
8	0.249	0.303

2.5.5 Independence of Irrelevant Alternatives (IIA)

The MLM is the most commonly used regression model when the dependent variable is categorical and the data structure is choice specific. The most important concern about the model, from an econometric perspective, is the assumption of the independence of irrelevant alternatives (IIA) (Cheng and Long, 2007). According to this assumption, the outcome categories of the model should not be affected if an 'irrelevant' alternative outcome category is added or deleted. For example, if the IIA assumption is not violated, then the relative probabilities of choosing academic or vocational education should not change if a choice of another type of education is added as an additional possibility.

The validity of the IIA assumption has been tested by computing the two most commonly used tests for the IIA assumption: Hausman-Mcfadden (HM) test (Hausman and McFadden, 1984) and the Small-Hsiao (SH) test (Small and Hsiao, 1985). The IIA tests compare the estimated coefficients from the full model with the restricted model which excludes at least one of the alternatives. HM test is asymptotically distributed as a chi^2 with degrees of freedom equal to the number of coefficients of the restricted model while the SH test is asymptotically distributed as a chi^2 with degrees of freedom being equal to the number of coefficients fitted in the full model as well as in the restricted model. A significant value of

chi^2, when $p < 0.01$, shows that the IIA assumption has been violated suggesting that the model is not acceptable. Although both tests are widely used by econometricians, it is worth mentioning that there is a general consciousness that both HM and SH tests in some cases provide conflicting results as to whether the IIA assumptions are violated, some of the tests reject the null hypothesis, whereas others do not. In addition, simulation studies have shown that even in large sample data sets IIA tests may perform poorly as they often reject the assumption when the alternatives seem distinct and often fail to reject IIA when the alternatives can reasonably be viewed as close substitutes (Fry and Harris, 1996, 1998, Cheng and Long, 2007).

The results from the HM and SH tests are reported in Table 2.13 for BCS and Table 2.14 for LSYPE. The findings from the HM test are mixed and cannot lead to valid conclusions about the assumption as, especially in the LSYPE sample, in most cases the HM test did not work. In the cases that the tests did work, both for BCS and LSYPE in some cases it did not reject the Ho that the IIA assumption holds while in other cases the Ho has been rejected. The SH tests (mostly) did not reject the null hypothesis (Ho) that the IIA assumption holds for all the estimated models using BCS and LSYPE. In general, the HM tests cannot lead to accurate conclusions about the validity of the models while the SH tests can lead to the conclusion that the IIA holds. At this point, it is worth mentioning that since the alternative options are sufficiently distinct from each other given the clear differences between staying and leaving full time education and conditional on staying selecting between an academic and a vocational route as well as the fact that there no additional alternatives, the statistical tests to provide evidence for independence of irrelevant alternatives are not really needed to justify this assumption. Economic theory provides enough justification with regards to this matter.

Table 2.13: BCS tests of IIA assumption

	Chi^2	df	Prob>chi^2	Evidence
Hausman tests				
Model 1				
Academic Education	5.796	21	1.000	for Ho
Vocational Education	291.455	22	0.000	against Ho
No post-compulsory Education	20.321	22	0.563	for Ho
Model 2				
Academic Education	262.653	16	0.000	against Ho
Vocational Education	144.608	16	0.000	against Ho
No post-compulsory Education	21.964	16	0.144	for Ho
Model 3				
Academic Education	2.445	22	1.000	for Ho
Vocational Education	356.628	22	0.000	against Ho
No post-compulsory Education	24.347	22	0.329	for Ho
Model 4				
Academic Education	3.030	16	1.000	for Ho
Vocational Education	351.064	16	0.000	against Ho
No post-compulsory Education	35.874	16	0.003	against Ho
Small-Hsiao tests				
Model 1				
Academic Education	35.563	22	0.034	against Ho
Vocational Education	37.658	22	0.020	against Ho
No post-compulsory Education	12.208	22	0.953	for Ho
Model 2				
Academic Education	12.775	16	0.689	for Ho
Vocational Education	11.371	16	0.786	for Ho
No post-compulsory Education	13.746	16	0.618	for Ho
Model 3				
Academic Education	16.2555	22	0.83	for Ho
Vocational Education	17.527	22	0.734	for Ho
No post-compulsory Education	19.218	22	0.632	for Ho
Model 4				
Academic Education	15.626	16	0.479	for Ho
Vocational Education	14.496	16	0.562	for Ho
No post-compulsory Education	16.021	16	0.452	for Ho

Table 2.14: LSYPE tests of IIA assumption

	$\mathbf{Chi^2}$	**df**	**Prob>chi^2**	**Evidence**
Hausman tests				
Model 5				
Academic Education	-0.684	6	.	.
Vocational Education	7.408	8	0.493	for Ho
No post-compulsory Education	8.945	22	0.994	for Ho
Model 6				
Academic Education	-1.429	15	.	.
Vocational Education	7.498	16	0.962	for Ho
No post-compulsory Education	-11.128	16	.	.
Model 7				
Academic Education	-1.240	21	.	.
Vocational Education	47.455	22	0.001	against Ho
No post-compulsory Education	34.036	22	0.049	for Ho
Model 8				
Academic Education	-1.804	15	.	.
Vocational Education	-146.835	16	.	.
No post-compulsory Education	4.117	16	0.999	for Ho
Small-Hsiao tests				
Model 5				
Academic Education	31.210	22	0.092	for Ho
Vocational Education	23.298	22	0.385	for Ho
No post-compulsory Education	27.004	22	0.2100	for Ho
Model 6				
Academic Education	25.324	16	0.064	for Ho
Vocational Education	24.211	16	0.085	for Ho
No post-compulsory Education	20.627	16	0.193	for Ho
Model 7				
Academic Education	25.049	22	0.295	for Ho
Vocational Education	26.307	22	0.239	for Ho
No post-compulsory Education	24.751	22	0.309	for Ho
Model 8				
Academic Education	11.146	16	0.800	for Ho
Vocational Education	18.886	16	0.276	for Ho
No post-compulsory Education	22.100	16	0.140	for Ho

2.6 Conclusion

This chapter provides empirical evidence on the influence of socio-economic background and ability on educational aspirations and choices as estimated in two different time periods, two different ages and using two different socio-economic components. The hypothesis of this chapter was whether the role of the socio-economic background on educational aspirations and choices increased through the years while the role of ability decreased. The study used data from two different panel surveys capturing the effect at two separate periods in time. Using the MLM, marginal effects have been computed, reported and analysed in depth for all cases.

The chapter presents evidence of a weakening link between both ability and socio-economic background in explaining educational aspirations for academic and no post-compulsory education which remains significantly decreased for final choices. The expansion of academic education which is identified in the recent cohort, both in terms of aspirations and actual participation, suggests a decline of any possible socio-economic influences that could be limiting attendance in the past. This research shows that the expansion of the education system has proportionately benefited individuals from all social-backgrounds and increased the chances for all students to acquire an academic qualification. On the other hand, the decreasing importance of ability suggests that there is an increase in the attainment of the least able students to academic education and therefore that it is very likely that the standards have fallen.

This chapter appears to partly agree and partly contradict the findings of Galindo-Rueda and Vignoles (2005b) who compared the BCS sample with an older sample and showed evidence of decreasing ability but increasing socio-economic effects, although they did not include vocational education in their estimated models. On the other hand, these findings appear to agree with many of the theoretical literature, which suggested that improved educational opportunities for disadvantaged students would lead to less educational inequality (Benabou, 1996, Fernandez and Rogerson, 1996, 1998). The extent at which this expansion is beneficial for the English labour market is still to be questioned.

The findings of this research should raise concerns to policy-makers that a further expansion of academic education is likely to cause more damage to the rest of the educational system and eventually lead to an unbalanced labour market with an oversupply of graduate individuals who are likely to take occupational positions or earnings that are inappropriate

for their skills. The estimations on vocational education, although instructive and of significant importance can be also described as disappointing as these findings reveal that vocational education in England follows a downward path. Vocational education should be considered highly important and necessary in a labour market where there remain many jobs with below degree level requirements.

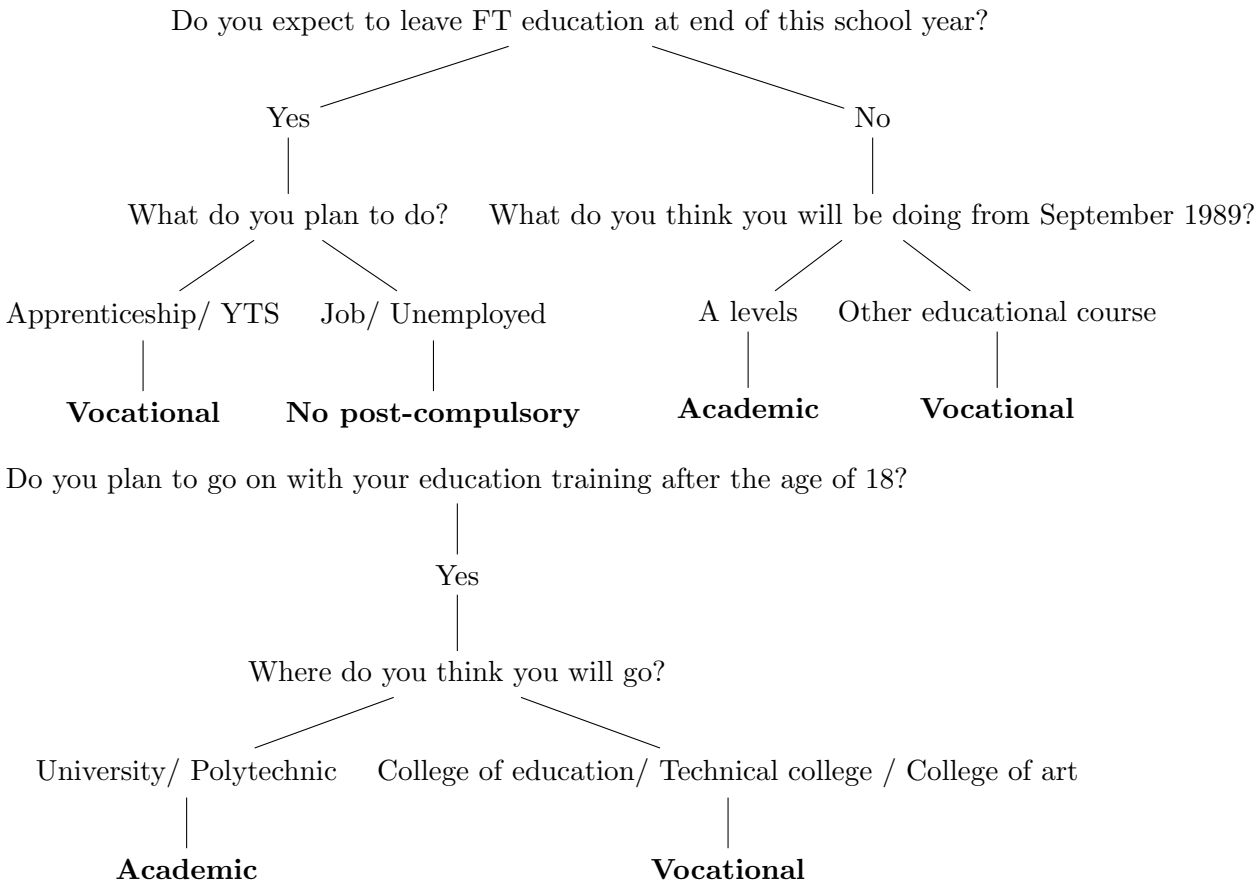
This research identifies a poor participation in vocational education in the more recent cohort which is significantly lower than in the older cohort. It also identifies increasing socio-economic and ability effects on vocational aspirations implying that the importance of vocational education is not recognised from the English population. In the older cohort, there has been identified a significant participation in vocational education which declined in the recent cohort. Looking though at the correlation of aspirations and choices, the low aspirations for vocational education imply that the individuals might have been 'forced' to that decision rather than being their direct goal from the beginning. The findings of this chapter suggest that the limited appreciation for vocational education is not a phenomenon that recently appeared. Lower ability individuals with parents in lower social-classes, both in terms of education and occupation, are found to be the ones who are most likely to follow vocational education. Dearden et al. (2009) mentioned that the higher wage premium that exists in the British labour market for academic qualifications provides evidence that vocational qualifications are not valued by employers. The analysis of educational aspirations for vocational qualifications from both cohorts reveal a push away from vocational training and suggest that it is the value of vocational education which needs to be reformed and recognised in order to make young individuals aspire to follow it.

Appendix A:

A.1 Derivation of educational aspirations variables

British Cohort Study

BCS 1986: Individuals' Questionnaire



Source: BCS1986, Student self-completion questionnaire (Document J), pp. 10-11

Observations: 4046

BCS 1980: Parents' Questionnaire

Question 1

At what age do you think your child will finally leave school?

Answer: 16 years old → **No post-compulsory**

Other answers: 17 years old, 18 years old

Question 2

Do you intend your child to continue his/her training after leaving school? If yes, what kind of education or training your child will have?

Answers:

College/ University → **Academic**

Apprenticeship → **Vocational**

Specific job → **Vocational**

Other answers: Further Education, Don't know

BCS 1986: Parents' Questionnaire

Question 3

Which of the following would you like your teenager to do and what do you think he/she will actually do after this school year?

Answers:

Leave at the end of this term → **No post-compulsory**

Stay in FT education and do vocational training → **Vocational**

Stay in FT education and do A levels → **Academic**

Other answers: Continue some form of FT education beyond age of 18, Other

Sources:

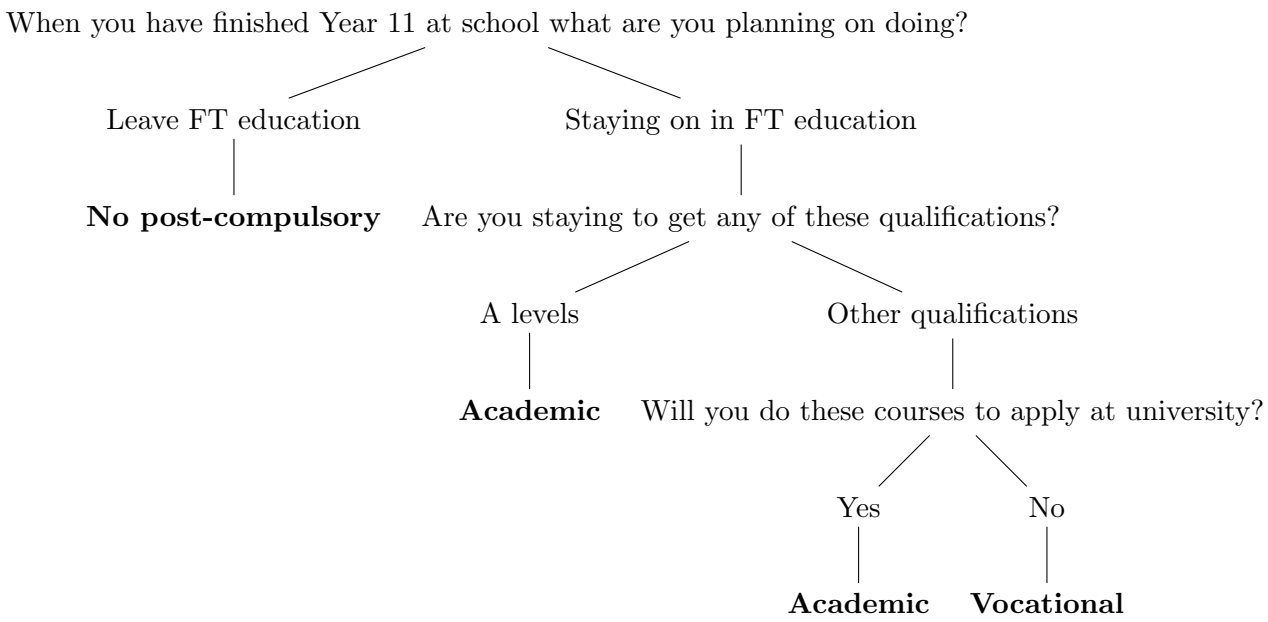
BCS1980, Maternal self-completion form, Section B: The child at school, pp. 7

BCS1986, Maternal self-completion form, Section B: The School, pp. 5

Observations: 5642

Longitudinal Study of Young People in England

LSYPE W3 (2006): Individual's Questionnaire



Have you applied or are you planning to apply for an apprenticeship or training place?

Yes → **Vocational**

Source: LSYPE W3, Young Person section, Future plans and advice, pp. 22-25, pp. 32

Observations: 11737

A.2 Proxying individuals' aspirations with parents' aspirations for the BCS cohort

In order to handle missing data, which was particularly a problem for the aspirations variable of the BCS sample, the individuals' aspirations have been proxied from their parents' aspirations. Appendix A.1 presented all the details of the questions taken from the parents' questionnaire. Missing information on individuals' aspirations at age 16 have been proxied from parents' aspirations at age 16. Further missing aspirations have been proxied from parents' aspirations at age 10 when information was available. To further validate this method the similarity of parents' and children's aspirations, for those observations which both could be observed, have also been examined. The aspirations of both parents and children could be observed from the 16-year follow-up for 3540 individuals and the correlation found between the two is presented in Table A.2.1 from the cross-tabulation of the two variables. It is evident from the available data that although the correlation between the two is not perfect, it is close enough to consider parent's aspirations as a good proxy for the individuals' aspirations.

Table A.2.1: Correlation of individuals' and parents' aspirations

	Individuals' Aspirations			
Parents' Aspirations	**Academic Education**	**Vocational Education**	**No post-comp. Education**	**Total**
Academic Education	1355 (0.74)	284 (0.16)	188 (0.10)	1827 (1)
Vocational Education	43 (0.08)	396 (0.70)	129 (0.23)	568 (1)
No post-comp. Education	48 (0.04)	514 (0.45)	583 (0.51)	1145 (1)
Total	1446 (0.41)	1194 (0.34)	900 (0.25)	3540 (1)

A.3 Derivation of educational choices variables

Appendix A.3 illustrates the derivation procedure of the educational choices variables for the BCS and LSYPE datasets respectively, presenting in which educational alternative each qualification is included.

Table A.3.1: BCS: Derivation of educational choices variable

	Academic Qualification	Vocational Qualification	No post-comp. Qualification
Age 26 follow-up:			
GCSE/CSE			✓
O Levels			✓
Scottish O Grade			✓
Scottish Standard Grade			✓
A/AS level	✓		
Scottish Higher Grade	✓		
Scottish Certificate of 6th form studies	✓		
HE Diploma	✓		
Degree (BA, Bsc, Bed, etc)	✓		
PGCE	✓		
Post Graduate Certificate	✓		
Post Graduate Degree (MA, Msc, PhD, etc)	✓		
HE Foundation course	✓		
Other HE qualification	✓		
NVQs		✓	
RSA		✓	
Pitmans		✓	
City and Guilds		✓	
JIB/NJC		✓	
ONC/ OND HNC/HND		✓	
TEC/ BEC/ BTEC		✓	
SCOTEC/ SCOTBEC/ SCOTVEC		✓	
Technical or business Qualification		✓	
Professional Qualification		✓	
Nursing Qualification		✓	
Age 29 follow-up:			
None academic qualification			✓
Bad GCSEs			✓
CSE 2-5, other Scottish school qualification			✓
O levels, Good GCSEs			✓
1 A level or more than 1 AS level	✓		
2 or more A levels	✓		
Diploma of HE	✓		
Degree, other degree level	✓		
Higher degree	✓		
None vocational qualification			✓
NVQ Level 1		✓	
NVQ Level 2		✓	
NVQ Level 3		✓	
NVQ Level 4		✓	
NVQ Level 5		✓	

Table A.3.2: LSYPE: Derivation of educational choices variable

	Academic Qualification	Vocational Qualification	No post-comp. Qualification
W4- W7: Academic qualification studying			
A/ AS levels	✓		
Applied A levels	✓		
Vocational A levels		✓	
W6: Highest academic qualification			
First/ Other Degree	✓		
Other HE	✓		
2+ A/AS	✓		
1 A/AS	✓		
W6: Whether currently at university			
Yes	✓		
W6-W7: Combined HE Qualification studied			
Degree	✓		
Foundation Degree	✓		
Teacher Training (BEd or BA/Bsc with QTs)	✓		
Diploma in HE	✓		
HND/HNC/RSA or OCR Higher Diploma/NVQ		✓	
W6-W7: HE Flag			
In HE	✓		
Accepted HE offer to start in 2009/10 or 2010/11	✓		
Applied for HE to start in 2009/10 or 2010/11	✓		
W7: Doing A levels to apply to university			
Yes	✓		
W4-W7: Studying vocational qualifications:			
Key Skills		✓	
Basic Skills		✓	
Foundation or intermediate GNVQs		✓	
NVQs		✓	
Edexcel, BTEC or LQL		✓	
OCR		✓	
City and Guilds		✓	
W4-W5: Other qualification is vocational			
Yes		✓	
W6-W7: Highest full NVQ qualification			
Level 1-5		✓	
Other HE		✓	
Level unknown		✓	
W7: Apprenticeship Flag			
Been in an Apprenticeship		✓	
W8: Academic Qualifications gained			
University Higher Degree	✓		
First degree level qualification	✓		
Diploma in HE	✓		
Teaching qualification (not PGCE)		✓	
Nursing qualification		✓	
A/ AS level	✓		
Welsh/ International Baccalaureate	✓		
Higher Grade/ Advance Higher (Scotland)	✓		
Certificate of sixth year studies	✓		
GCSE			✓

Table A.3.2 –continued on next page

Table A.3.2 – continued from previous page

Standard Grade/ Lower (Scotland)		✓
W8: Vocational Qualifications gained		
Youth training certificate	✓	
Key Skills	✓	
Basic Skills	✓	
Modern/ trade apprenticeship	✓	
RSA/ OCR / Clerical qualifications	✓	
City and Guilds	✓	
GNVQ/ GSVQ	✓	
NVQ/ SVQ	✓	
HNC/ HND	✓	
ONC/ OND	✓	
BTEC /BEC/ TEC/ EdExcel /LQL	✓	
SCOTVEC, SCOTEC or SCOTBEC	✓	
Other vocational, technical or professional	✓	

A.4 Qualification check-list

Table A.4.1: Qualification check-list

Qualification	BCS	LSYPE
Academic		
Post-graduate degree (MA, Msc, PhD, PGCE)	✓	✓
Post-graduate certificate	✓	
Degree (BA, Bsc, Bed, etc)	✓	✓
HE Diploma	✓	✓
Other HE qualification	✓	✓
A levels/ A2s / AS	✓	✓
Scottish Higher Grade	✓	✓
Certificate of sixth year studies	✓	✓
Foundation course	✓	
International/ Welsh Baccalaureate		✓
Vocational		
Teaching qualification	✓	✓
Nursing qualification	✓	✓
Professional qualification	✓	✓
Technical or business qualification	✓	✓
NVQ/ GNVQ	✓	✓
RSA/ OCR	✓	✓
City and Guilds	✓	✓
Pitmans	✓	✓
JIB/NJC	✓	
HNC/ HND	✓	✓
ONC/ OND	✓	✓
BTEC /BEC/ TEC/	✓	✓
Apprenticeship	✓	✓
Youth training certificate	✓	✓
Key Skills		✓
Basic Skills		✓
EdExcel /LQL		✓
Other vocational	✓	✓
No post-compulsory		
No qualifications	✓	✓
GCSE/ CSE	✓	✓
Standard Grade/ Lower (Scotland)	✓	✓

A.5 Description of control variables

Appendix A.5 provides all the necessary information concerning the control variables included in this chapter. There is detailed information explaining how each control variable has been measured and the source and time period of the data used for each variable. Some variables which had to be constructed under some special treatments, mainly to handle missing data problems, are also explained. For the dummy variables the first category mentioned has been used as the reference category of each variable.

Gender

BCS

Source: Birth survey (1970)

Type: Dummy

Measurement: Girl, Boy

LSYPE

Source: W1: 13-14 years old (2004)

Type: Dummy

Measurement: Girl, Boy

Ethnicity

BCS

Source: Birth Survey (1970)

Type: Dummy

Measurement: White, Other

LSYPE

Source: W1: 13-14 years old (2004)

Type: Dummy

Measurement: White, Other

Geographic Area (urban/rural indicator)

BCS

Source: 5-year follow-up (1975)

Type: Dummy

Measurement: Urban, Rural

LSYPE

Source: W3: 15-16 years old (2006)

Type: Dummy

Measurement: Urban, Rural

Parents' age

BCS

Source: 5-year follow-up (1975)

Type: Continuous

Description: Parent's age at child's birth

Treatment: The missing observations have been replaced with the average age

LSYPE

Source: W1: 13-14 years old (2004)

Type: Continuous

Description: Parents' age at W1

Treatment: The missing observations have been replaced with the average age

Number of children in the household

BCS

Source: 10-year follow-up (1980)

Type: Continuous

Description: Total number of children in each household including the individual

Measurement: From 1 (no siblings) up to 11

LSYPE

Source: W1: 13-14 years old (2004)

Type: Continuous

Description: Total number of children in each household including the individual

Measurement: From 1 (no siblings) up to 14

A.6 Descriptive statistics

Table A.6.1: Descriptive Statistics

	BCS (1970)		**LSYPE (1990)**	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Educational Aspirations				
Academic	0.34	0.47	0.74	0.44
Vocational	0.25	0.43	0.23	0.42
No post-compulsory	0.41	0.49	0.04	0.19
Educational Choices				
Academic	0.31	0.46	0.67	0.47
Vocational	0.49	0.50	0.25	0.43
No post-compulsory	0.20	0.40	0.08	0.28
Individual ability	0.10	0.99	0.13	0.97
Parents' Occupation				
Father Professional	0.06	0.23	0.04	0.20
Father Managerial	0.22	0.42	0.25	0.43
Father Skilled-Manual	0.08	0.27	0.06	0.24
Father Skilled Non-Manual	0.39	0.49	0.24	0.43
Father Partly Skilled	0.10	0.31	0.06	0.24
Father Unskilled	0.03	0.17	0.02	0.14
Father Unemployed	0.01	0.10	0.02	0.15
Father Missing	0.11	0.32	0.30	0.46
Mother Professional	0.00	0.06	0.02	0.13
Mother Managerial	0.12	0.33	0.24	0.43
Mother Skilled-Manual	0.23	0.42	0.26	0.44
Mother Skilled Non-Manual	0.07	0.25	0.12	0.32
Mother Partly Skilled	0.20	0.40	0.18	0.39
Mother Unskilled	0.06	0.25	0.03	0.18
Mother Unemployed	0.23	0.42	0.12	0.32
Mother Missing	0.08	0.27	0.03	0.17
Parents' Education				
Father Academic Degree	0.13	0.34	0.18	0.38
Father A levels	0.09	0.28	0.12	0.33
Father Other Qualification	0.34	0.48	0.23	0.42
Father No qualification	0.32	0.47	0.16	0.37
Father Missing	0.12	0.32	0.31	0.46
Mother Academic Degree	0.06	0.23	0.23	0.42
Mother A levels	0.04	0.19	0.13	0.33
Mother Other Qualification	0.35	0.48	0.38	0.49
Mother No qualification	0.47	0.50	0.23	0.42
Mother Missing	0.08	0.28	0.03	0.18
Control Variables				
Girl	0.52	0.50	0.50	0.50
White	0.97	0.17	0.71	0.45
Father's Age	28.67	5.84	44.65	5.98
Mother's Age	26.02	5.30	41.65	5.43
Children in household	2.64	1.08	2.98	1.44
Urban	0.88	0.33	0.82	0.38
Observations	6652		9405	

A.7 Comparison of total and analytical sample

Table A.7.1: BCS: Comparison of total and analytical sample

	Total Sample			Analytical Sample			
Variable	**N**	**Mean**	**Std. Dev.**	**N**	**Mean**	**Std. Dev.**	**Difference**
Educational Aspirations							
Academic	9796	0.32	0.47	6652	0.34	0.47	-0.02(0.01)**
Vocational	9796	0.23	0.42	6652	0.25	0.43	-0.02(0.01)***
No post-compulsory	9796	0.44	0.50	6652	0.41	0.49	0.03(0.01)***
Educational Choices							
Academic	9400	0.30	0.46	6652	0.31	0.46	-0.01(0.01)*
Vocational	9400	0.48	0.50	6652	0.49	0.50	-0.01(0.01)
No post-compulsory	9400	0.22	0.41	6652	0.20	0.40	0.02(0.01)**
Individual ability	10282	0	1	6652	0.10	0.99	-0.10(0.01)***
Parents' Occupation							
Father Professional	12118	0.05	0.22	6652	0.06	0.23	-0.01(0.00)**
Father Managerial	12118	0.20	0.40	6652	0.22	0.42	-0.02(0.01)***
Father Skilled-Manual	12118	0.07	0.26	6652	0.08	0.27	-0.01(0.00)
Father Skilled Non-Manual	12118	0.37	0.48	6652	0.39	0.49	-0.02(0.01)**
Father Partly Skilled	12118	0.10	0.30	6652	0.10	0.31	-0.00(0.00)
Father Unskilled	12118	0.03	0.18	6652	0.03	0.17	0.00(0.00)*
Father Unemployed	12118	0.01	0.11	6652	0.01	0.10	0.00(0.00)
Father Missing	12118	0.16	0.37	6652	0.11	0.32	0.05(0.01)***
Mother Professional	12118	0.00	0.05	6652	0.00	0.06	-0.00(0.00)
Mother Managerial	12118	0.11	0.32	6652	0.12	0.33	-0.01(0.00)**
Mother Skilled-Manual	12118	0.22	0.41	6652	0.23	0.42	-0.01(0.01)***
Mother Skilled Non-Manual	12118	0.06	0.24	6652	0.07	0.25	-0.01(0.00)
Mother Partly Skilled	12118	0.19	0.39	6652	0.20	0.40	-0.01(0.00)
Mother Unskilled	12118	0.06	0.24	6652	0.06	0.25	-0.00(0.00)
Mother Unemployed	12118	0.23	0.42	6652	0.23	0.42	-0.00(0.01)
Mother Missing	12118	0.12	0.33	6652	0.08	0.27	0.04(0.00)***
Parents' Education							
Father Academic Degree	12118	0.12	0.32	6652	0.13	0.34	-0.01(0.00)***
Father A levels	12118	0.08	0.27	6652	0.09	0.28	-0.01(0.00)*
Father Other Qualification	12118	0.32	0.47	6652	0.34	0.48	-0.02(0.01)***
Father No qualification	12118	0.32	0.47	6652	0.32	0.47	-0.00(0.00)
Father Missing	12118	0.17	0.37	6652	0.12	0.32	0.05(0.01)***
Mother Academic Degree	12118	0.05	0.21	6652	0.06	0.23	-0.01(0.00)***
Mother A levels	12118	0.04	0.19	6652	0.04	0.19	-0.00(0.00)
Mother Other Qualification	12118	0.33	0.47	6652	0.35	0.48	-0.02(0.01)***
Mother No qualification	12118	0.46	0.50	6652	0.47	0.50	-0.01(0.01)
Mother Missing	12118	0.13	0.34	6652	0.08	0.28	0.05(0.00)***
Control Variables							
Girl	12117	0.48	0.50	6652	0.52	0.50	-0.04(0.01)***
White	11901	0.96	0.20	6652	0.97	0.17	-0.01(0.00)***
Father's Age	12118	28.60	5.86	6652	28.67	5.84	-0.07(0.09)
Mother's Age	12118	25.97	5.43	6652	26.02	5.30	-0.05(0.08)
Children in household	11839	2.69	1.15	6652	2.64	1.08	0.05(0.02)***
Urban	12052	0.89	0.32	6652	0.88	0.33	0.01(0.00)**

The last column presents the t-test estimates for mean differences of each variable in the total and analytical sample.
Standard errors are presented in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7.2: LSYPE: Comparison of total and analytical sample

	Total Sample			Analytical Sample			
Variable	**N**	**Mean**	**Std. Dev.**	**N**	**Mean**	**Std. Dev.**	**Difference**
Educational Aspirations							
Academic	11737	0.72	0.45	9405	0.74	0.44	-0.02(0.01)**
Vocational	11737	0.24	0.43	9405	0.23	0.42	0.01(0.00)*
No post-compulsory	11737	0.04	0.19	9405	0.04	0.19	0.00(0.00)
Educational Choices							
Academic	12409	0.64	0.48	9405	0.67	0.47	-0.03(0.01)***
Vocational	12409	0.26	0.44	9405	0.25	0.43	0.01(0.01)**
No post-compulsory	12409	0.11	0.31	9405	0.08	0.28	0.03(0.00)***
Individual ability	14173	0	1	9405	0.13	0.97	-0.13(0.01)***
Parents' Occupation							
Father Professional	15770	0.04	0.19	9405	0.04	0.20	-0.00(0.00)*
Father Managerial	15770	0.21	0.41	9405	0.25	0.43	-0.04(0.01)***
Father Skilled-Manual	15770	0.05	0.22	9405	0.06	0.24	-0.01(0.00)**
Father Skilled Non-Manual	15770	0.23	0.42	9405	0.24	0.43	-0.01(0.01)***
Father Partly Skilled	15770	0.06	0.24	9405	0.06	0.24	-0.00(0.00)
Father Unskilled	15770	0.02	0.15	9405	0.02	0.14	0.00(0.00)
Father Unemployed	15770	0.03	0.17	9405	0.02	0.15	0.01(0.00)**
Father Missing	15770	0.36	0.48	9405	0.30	0.46	0.06(0.00)***
Mother Professional	15770	0.02	0.13	9405	0.02	0.13	-0.00(0.00)
Mother Managerial	15770	0.21	0.41	9405	0.24	0.43	-0.03(0.00)***
Mother Skilled-Manual	15770	0.23	0.42	9405	0.26	0.44	-0.03(0.00)***
Mother Skilled Non-Manual	15770	0.11	0.31	9405	0.12	0.32	-0.01(0.00)**
Mother Partly Skilled	15770	0.17	0.38	9405	0.18	0.39	-0.01(0.00)**
Mother Unskilled	15770	0.03	0.18	9405	0.03	0.18	0.00(0.00)
Mother Unemployed	15770	0.13	0.34	9405	0.12	0.32	0.01(0.00)***
Mother Missing	15770	0.10	0.29	9405	0.03	0.17	0.07(0.00)***
Parents' Education							
Father Academic Degree	15770	0.15	0.36	9405	0.18	0.38	-0.03(0.00)***
Father A levels	15770	0.11	0.31	9405	0.12	0.33	-0.01(0.00)***
Father Other Qualification	15770	0.21	0.41	9405	0.23	0.42	-0.02(0.00)***
Father No qualification	15770	0.17	0.37	9405	0.16	0.37	0.01(0.00)*
Father Missing	15770	0.37	0.48	9405	0.31	0.46	0.06(0.01)***
Mother Academic Degree	15770	0.20	0.40	9405	0.23	0.42	-0.03(0.00)***
Mother A levels	15770	0.11	0.31	9405	0.13	0.33	-0.02(0.00)***
Mother Other Qualification	15770	0.35	0.48	9405	0.38	0.49	-0.03(0.00)***
Mother No qualification	15770	0.25	0.43	9405	0.23	0.42	0.02(0.00)***
Mother Missing	15770	0.10	0.30	9405	0.03	0.18	0.07(0.00)***
Control Variables							
Girl	15431	0.49	0.50	9405	0.50	0.50	-0.01(0.00)
White	15744	0.67	0.47	9405	0.71	0.45	-0.04(0.00)***
Father's Age	15770	44.56	6.08	9405	44.65	5.98	-0.09(0.08)
Mother's Age	15219	41.34	5.60	9405	41.65	5.43	-0.31(0.07)***
Children in household	15068	3.06	1.53	9405	2.98	1.44	0.08(0.02)***
Urban	13531	0.84	0.37	9405	0.82	0.38	0.02(0.00)***

The last column presents the t-test estimates for mean differences of each variable in the total and analytical sample.
Standard errors are presented in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.8 BCS cohort's member questionnaire

Appendix A.8 shows the results of the estimations on educational aspirations of the BCS sample using only the cohort's member questionnaire from the 16-year follow-up. The estimations have been carried out as a robustness check of the main estimations.

Table A.8.1: Cohort member's Questionnaire: MLM Marginal Effects on educational aspirations measuring SES with parental occupation

	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No-post compulsory Education Mfx
Arithmetic test	0.223***	-0.138***	-0.084***
	(0.007)	(0.008)	(0.008)
Father Professional	0.335***	0.086	-0.421***
	(0.072)	(0.072)	(0.063)
Father Managerial	0.217***	0.062	-0.279***
	(0.068)	(0.062)	(0.046)
Father Skilled-Manual	0.189***	0.061	-0.249***
	(0.070)	(0.066)	(0.050)
Father Skilled Non-Manual	0.096	0.108*	-0.204***
	(0.068)	(0.061)	(0.044)
Father Partly Skilled	0.074	0.124*	-0.198***
	(0.071)	(0.064)	(0.048)
Father Unemployed	-0.078	0.178*	-0.099
	(0.124)	(0.107)	(0.085)
Father Missing	0.134*	0.078	-0.212***
	(0.070)	(0.065)	(0.049)
Mother Professional	0.890	1.223	-2.114
	(27.263)	(43.891)	(71.154)
Mother Managerial	0.149***	-0.044	-0.105***
	(0.041)	(0.042)	(0.040)
Mother Skilled-Manual	0.049	-0.055	0.006
	(0.039)	(0.037)	(0.033)
Mother Skilled Non-Manual	0.045	0.037	-0.082**
	(0.047)	(0.045)	(0.041)
Mother Partly Skilled	-0.009	-0.003	0.013
	(0.041)	(0.038)	(0.033)
Mother Unemployed	0.073*	-0.041	-0.032
	(0.039)	(0.037)	(0.033)
Mother Missing	0.073	-0.073	-0.000
	(0.045)	(0.045)	(0.040)
Girl	-0.002	0.055***	-0.053***
	(0.015)	(0.017)	(0.015)
White	-0.370***	0.147***	0.223***
	(0.046)	(0.056)	(0.057)
Father's Age	0.000	0.000	-0.001
	(0.001)	(0.001)	(0.001)
Mother's Age	0.004***	-0.002	-0.003
	(0.001)	(0.002)	(0.002)
Children in household	-0.039***	0.015*	0.023***
	(0.008)	(0.008)	(0.007)
Urban	0.006	0.023	-0.030
	(0.021)	(0.024)	(0.022)
Observations	3108	3108	3108

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Omitted groups: *Occupation:* Unskilled *Area:* Rural *Gender:* Boy *Ethnicity:* Other

Table A.8.2: Cohort member's Questionnaire: MLM Marginal Effects on educational aspirations measuring SES with parental education

	(1) Academic Education Mfx	(2) Vocational Education Mfx	(3) No-post compulsory Education Mfx
Arithmetic test	0.214***	-0.133***	-0.080***
	(0.007)	(0.008)	(0.008)
Father Academic Degree	0.190***	-0.083**	-0.107***
	(0.024)	(0.033)	(0.033)
Father A levels	0.077***	-0.008	-0.069**
	(0.026)	(0.031)	(0.030)
Father Other Qualification	0.033*	-0.017	-0.017
	(0.019)	(0.021)	(0.019)
Father Missing	0.032	-0.059*	0.028
	(0.029)	(0.031)	(0.028)
Mother Academic Degree	0.282***	0.024	-0.305***
	(0.038)	(0.061)	(0.075)
Mother A levels	0.197***	-0.037	-0.160***
	(0.036)	(0.049)	(0.051)
Mother Other Qualification	0.094***	-0.028	-0.066***
	(0.016)	(0.019)	(0.017)
Mother Missing	0.023	0.010	-0.033
	(0.030)	(0.033)	(0.030)
Girl	-0.002	0.056***	-0.054***
	(0.015)	(0.016)	(0.015)
White	-0.354***	0.127**	0.227***
	(0.044)	(0.056)	(0.058)
Father's Age	0.000	0.000	-0.001
	(0.001)	(0.001)	(0.001)
Mother's Age	0.004***	-0.001	-0.002
	(0.001)	(0.002)	(0.002)
Children in household	-0.036***	0.016*	0.020***
	(0.008)	(0.008)	(0.007)
Urban	0.005	0.019	-0.024
	(0.021)	(0.024)	(0.022)
Observations	3108	3108	3108

Standard errors in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Omitted groups: *Occupation:* Unskilled *Area:* Rural *Gender:* Boy *Ethnicity:* Other

Chapter 3

The causal effect of secondary school peers on educational aspirations[11]

3.1 Introduction

3.1.1 Background and motivation

The effects of peers on individuals' behaviour and outcomes have long been of interest in social sciences. Likewise, the factors that influence the education participation decisions of young people are of great importance for policy makers and other stakeholders, as well as being of concern to parents and the individuals themselves. The aim of this chapter is to bring these two areas together in order to investigate the influence of school peers on individuals' post-compulsory education participation decisions. In particular, the chapter examines the role that secondary school peers play in influencing the future educational aspirations of students before they complete their compulsory schooling.

The importance of peer effects in education arises from the multiplier effects that may be generated from student interactions. If a student's peers can have a causal impact on their outcomes, then decisions concerning class composition can affect the education production function and, in aggregate, potentially impact on macroeconomic growth (Sojourner, 2013, Hoxby, 2000). On the other hand, if peer-group quality impacts on an individual's achievements, then sorting across schools by prior ability could exacerbate educational inequalities, and consequently reinforce existing disadvantage.

Although not irreversible, the decisions that young people make regarding their future education tend to have significant effects on the rest of their lives. These will include both economic outcomes (such as wages and employment) and non-economic outcomes (for example, health). Much academic study has investigated the influences on educational participation decisions, with prior attainment and socio-economic background receiving most attention. Clearly, if peer influences are an important factor in determining outcomes,

[11] *An edited version of this chapter is published at CVER's Discussion Paper Series (http://cver.lse.ac.uk/textonly/cver/pubs/cverdp017.pdf).*

then parents, teachers and policy makers will all be interested in the magnitude, composition and determinants of these peer effects (Sacerdote, 2011).

According to Manski (1993), individuals belonging to the same peer group tend to behave similarly for three different reasons: 'endogenous effects', where the propensity of an individual to behave in some way varies with the behaviour of the group; 'exogenous effects', where the propensity of an individual to behave in some way varies with the exogenous characteristics of the group; and 'correlated effects' whereby individuals in the same group tend to behave similarly because they have similar characteristics or face similar institutional environments[12].

The baseline model for estimating peer effects is the linear-in-means model. The model associates individuals' outcomes with their own characteristics as well as the average characteristics of their peers. From an empirical point of view, the baseline model has two fundamental shortcomings in measuring peer effects. First, it is often difficult to separate the effect that the peer group has on the individual from the effect the individual has on the group. Because outcomes are simultaneous, an individual will affect their peers as much as their peers affect them (the so-called *'reflection problem'*). If the reflection problem is not taken into consideration, the linear-in-means model would be a biased estimator for peer effects. Second, peer groups are seldom randomly compiled. Rather, they are typically, at least to some degree, self-selected. This self-selection into peer groups can generate effects unobserved to the researcher which are correlated with peer characteristics. Thus in the presence of self-selection, it is difficult to distinguish peer effects from selection effects (de Xavier Pinto, 2010, Robertson and Symons, 2003). If individuals self-select into groups, then selection bias would arise from the fact that an outcome which appeared to be a peer effect is really just a consequence of the fact that people who act in a similar way or who share similar characteristics make themselves into groups. As Hoxby (2000) explained, if every individual in a group appears to be high achieving then selection bias could arise when an observer assumes that achievement is an effect of being in that group instead of a reason for being in it.

[12]In the context of this study, an endogenous effect arises if individuals aspirations vary with the average achievement of the peer group; an exogenous effect arises if the individuals' aspirations vary with the observable socio-economic characteristics of the peer group; and correlated effects arise if individuals have similar aspirations as their peers because they are subject to similar unobservable factors such as teacher quality.

3.1.2 Research question

This chapter uses the Longitudinal Study of Young People in England (LSYPE), which is the more recent cohort analysed in the previous chapter, to investigate the effect of secondary school peers on individuals' preliminary (Year 9, at age 14) and later (Year 11, at age 16) educational participation aspirations. Educational aspirations are defined, in this chapter, as the selection between leaving or staying in full-time education after completing the compulsory level and, conditional on staying, intentions to follow an academic or a vocational route. Peer influence is examined in two different ways; through peers' achievements, as estimated from average scores in KS3 exams and peers' own aspirations.

In order to eliminate selection bias associated with the choice of peer group the chapter adopts a novel identification strategy based on the peers-of-peers. Specifically, each individual's secondary school peers are instrumented with their primary school peers who did not attend the same primary and secondary school as the individual. The idea, as originally suggested by Mendolia et al. (2018), is that some of the peers of any specific secondary school student have had primary school peers who have never been directly exposed to the individual of interest because they went to a different primary and secondary school. Therefore, these peers-of-peers could not have a direct effect on the individual's aspirations.

3.1.3 Research findings and limitations

The main findings of this research are summarised as follows:

- Peer effects on individuals' aspirations to stay in education after age 16 are insignificant for girls but not for boys.
- Peers' ability and aspirations to follow an academic rather than a vocational route, have a positive and significant effect on both girls' and boys' aspirations to follow an academic pathway.
- The provision of information, advice and guidance (IAG) by school teachers and, in particular, by external agencies can serve to mitigate any peer effects - IAG appears to substitute for the influence of peers on individuals' aspirations.
- Individuals with higher ability peers are less likely to have changed their aspirations

between Year 9 and Year 11 (final year) of compulsory schooling.

A possible limitation of this chapter as a consequence of data availability is possibly the specification of the peer group, consisting of all students in the same grade and secondary school, which could be described as been specified at a rather broad level. As explained by Lin (2010) each student is not equally affected by all other students in the same school or grade but instead, they are more likely to be significantly influenced by some of them, such as their friends. An ideal model for estimating peer effects would be one containing the weighted average of the peer variables, with weight determined by the importance of a friend, as opposed to a mean peer variable. Of course, such data are rarely available. Further, potential threads to identification of the causal peer effect due to limitation in the construction of the instrumental variables used in the statistical analysis of this chapter are discussed later on together with potential solutions for improvement in the future.

3.1.4 Structure

The remainder of this chapter is structured as follows. In the next section there is a comprehensive review of the relevant literature on peer effects in education. Section 3.3 describes the data and explains the peer-effect indicators and outcomes. Section 3.4 explains the identification strategy that has been used to estimate the causal effect of secondary school peers on educational aspirations and Section 3.5 presents the main empirical findings. Finally, Section 3.6 concludes this chapter.

3.2 Literature Review

3.2.1 Peer effects and the identification problem

As defined by Gibbons and Telhaj (2016), peer effects are a distinct class of influences arising from social interactions, a broad term which encompasses any type of individual behaviour that involves interdependency with the behaviour or characteristics of others. In the literature, there is a general consensus concerning the correlation of an individual's outcome to that of their peers, however, the extent to which this effect is causal is the subject of extensive research (Goux and Maurin, 2007). Peer effects have been estimated for various outcomes including criminal behaviour, alcohol and drug consumption, smoking, pregnancy, obesity, sexual behaviour, retirement, charitable giving and more commonly

education.

Generally, it is difficult for researchers to determine whether they are observing peer effects or simply observing similar people behaving similarly (Winston and Zimmerman, 2004). Gibbons and Telhaj (2016) and Cooley (2014) acknowledged that measurements of peer-group characteristics may be very good proxies for unobserved individual, family background or institutional factors that can affect student attainment, making peer effects look important when they might not be. As Goldsmith-Pinkham and Imbens (2013) explained, the specific concern is that individuals have common characteristics that are correlated with their outcomes and that these characteristics also affect the formation of links. Individuals tend to exhibit homophile in these unobserved characteristics making it more likely that individuals who share similar values for these characteristics to form links. If these characteristics are also correlated with their outcomes, then researchers will find that individuals who are connected have correlated outcomes even though there are not peer effects.

While the empirical evidence on peer effects is growing, still, the results of the various studies provide mixed evidence regarding the magnitude and even sign of the peer effect which reflects the difficulty in defining the peer-group, isolating causal peer-group effects from other influences, lack of appropriate data and different identification methodologies adopted by researchers (Gibbons and Telhaj, 2016). The literature exploring peer effects in non-educational outcomes mostly identifies larger peer effects and the existence of the effects is possibly less controversial than that on the educational literature (Sacerdote, 2011). Studies exploring peer effects on educational outcomes have a considerable disagreement as to the sign and magnitude of the peer effect and also several authors find that peer effects disappear or become nearly insignificant when appropriate econometric techniques are implemented or once an individual's background characteristics are controlled for (Cohen-Cole and Fletcher, 2008, Mayer and Jencks, 1989).

3.2.2 Peer effects in education

Introduction

Quantifying externalities in education hold the attention of policy makers, schools, parents and teachers interested in efficient educational production given student heterogeneity (Foster, 2003). Peer effects on own educational outcomes constitute one form of these

educational externalities and have been extensively explored by applied microeconomists. The literature on peer effects in education dates back to the 1960s with the publication of the influential Coleman Report which persuaded researchers of the importance of peer effects in education (Coleman, 1968).

As the literature on education grows, pursuing a deeper understanding of the educational production function, many researchers and teachers have argued that peer composition is as important determinant of student outcomes as other widely cited inputs such as teacher quality, class size, family background and parental involvement. An important step in being able to identify causal peer effects is to remove or control for the selection of students into peer groups. Based on the pioneering work of Manski (1993), there is a growing literature that proposes alternative methods to estimate peer effects in education that overcome the shortcomings and challenges associated with identifying the parameters of the linear-in-means model. Social scientists have provided credible measurement and identification on the nature and size of peer effects using various econometric techniques to overcome selection in peer groups which would lead to biased estimations (de Xavier Pinto, 2010, Cooley, 2014).

The selection problem is typically handled using three main identification strategies. The first strategy relies on identifying some form of exogenous variation in the assignment of students and it is used by a growing literature measuring peer effects in Higher Education exploiting situations in which individuals are randomly assigned to university dorms and consequently to peer groups. The second strategy, widely used in the primary education peer-effect literature, is to exploit variation across classrooms or cohorts within a school while employing individual, school, teacher, grade, and year fixed effects. The fixed effects methodology relies on controlling the nearly inevitable self-selection of students into schools, classrooms and peer groups. The final strategy involves finding suitable instruments for peer behaviour that are exogenous to the stochastic error component of the dependent variable. The instrumental variable approach has become more popular in most recent studies on peer effects and relies on the fact that it affects the outcome only through its effect on the endogenous variable (Von Hinke et al., 2019, Carrell et al., 2009).

Empirical studies on peer effects, even when adopting the finest econometric techniques, rely on the assumption that peer spillovers can be measured through observables. Cooley (2014) discussed that, in the education context, many theories of peer spillovers center around

unobservables, such as ability, effort or motivation and showed that when peer effects arise from unobservables, the typical empirical specifications do not measure peer effects accurately. The findings of Cooley (2014) help explain the differences in the magnitude and even sign of peer effect estimates across the various studies.

The rest of the literature review on peer effects in education is structured as follows. The next part provides a detailed analysis of the relevant literature on peer effects in education employing the three main methodologies introduced above to make causal estimations. The reviewed literature includes the recent studies on peer effects in education, particularly the relevant literature of peer effects on educational achievements since 2000. Although the outcome of interest across the majority of the literature has been almost exclusively educational achievements, the definition of the peer group differs across the various studies. The reviewed literature covers peer effects on educational achievements in both school and college level covering various peer groups including, friends, schoolmates, classmates, studymates, fellow students and roommates or dormmates. Following, the next part discusses peer effects in education beyond achievements, examining outcomes other than test or exam scores. Finally, the last part summarises the key studies on peer effects in the UK, reviewing the main data sources and the extent that the existing literature instructs about the existence of peer effects in British schools.

Peer effects on achievements

Random Assignment

The random assignment methodology has been adopted by various recent papers examining peer effects in an attempt to overcome the identification problem. The specific methodology has been mainly, though not exclusively, adopted by researchers examining peer effects in Higher Education utilising the random assignment of students into housing units in order to examine the effects of roommates, dormmates and squadron members on students' achievements.

The papers of Sacerdote (2001) and Zimmerman (2003) which examined the importance of peer effects using college roommates who had been randomly assigned to dorms have received considerable attention mainly due to the general difficulty of finding credible exogenous variation in peer quality. Both studies found positive peer effects on a student's first year grade performance, the evidence they provide, though, is limited and not robust to sample modifications or alternative specifications. Specifically, Sacerdote (2001) found

no evidence that students' first year grade point average is influenced by their roommates' score on an academic index created by the Dartmouth admissions office if this score is included in the specified grade regression linearly. In a different specification, though, having a roommate with an academic index score in the top 25% is found to increase a student's grade point average by .033 points relative to having a roommate with a score in the bottom 25% and by .047 points relative to having a roommate with a score in the middle 50%. Zimmerman (2003) found no evidence that students' first year grade point average is influenced by their roommates' total Scholastic Aptitude Test (SAT) score but found evidence that their first year grades are positively correlated with their roommates' verbal SAT score if the roommate's Maths SAT score is also included in the regression specification.

The most recent works of Stinebrickner and Stinebrickner (2003) and Foster (2003) suggested that the mixed results of the previous studies could either reflect that peer effects do not play a particularly important role in Higher Education or that these empirical efforts did not look in the"right place" to find the evidence. Specifically, Stinebrickner and Stinebrickner (2003) suggested that it is very likely that the high selectivity of both Dartmouth College and Williams College made it unclear that these studies have been looking at the performance of the types of students who would benefit substantially from peers. Stinebrickner and Stinebrickner (2003) also argued that while policy interest in peer effects typically arises in contexts where some of the students of interest are of low ability or are from disadvantaged backgrounds, virtually all students at Dartmouth and Williams Colleges are of very high quality which substantially mitigates the potential influence of peer effects. In addition, Foster (2003) discussed that Zimmerman (2003) interpreted his mixed results as supporting a very specific set of true reduced-form functions of peer effects rather than evidence of lack of robustness for peer effects. While both the study of Stinebrickner and Stinebrickner (2003) and Foster (2003) argue to use data which better represent the U.S. college population, than data used in previous studies, still their findings are not equivalent. Stinebrickner and Stinebrickner (2003) found evidence of positive peer effects only for girls while found no significant peer effect for boys indicating differences in the importance of peer effects by gender. Foster (2003) found that conventional peer effects are insignificant and not robust in her sample.

Winston and Zimmerman (2004) adopted a similar approach exploring peer effects in Higher Education achievements of randomly assigned roommates for three different schools.

Although their evidence of peer effects was mixed and did not remain positive for all three schools the authors concluded that they are confident to say that peer effects in Higher Education do exist and that the signs of those effects are in the direction that would indicate institutional selectivity-strong students tend to increase peers' academic performance, and weak students tend to reduce it. Further, Hoel et al. (2005) studied peer effects in academic performance using randomly assigned students from Reed College in classrooms, rooms and dorms. The authors found no significant classroom peer effects but identified robust roommates and dormmates effects on student performance both contemporaneously in the first year and even extending through the students' undergraduate career. Similarly, McEwan and Soderberg (2006) using data from Wellesley College, explored peer effects taking advantage of the random allocation of first year students to their roommates. The authors' findings are mostly consistent with the previous studies finding suggestive evidence of non-linearities which, though, are not robust across alternate specifications.

Evidently, there is a considerable literature exploring peer effects using the random assignment technique but still little evidence of large positive peer effects in academic performance is found. Carrell et al. (2009) suggested that a major drawback of these studies is that roommates are generally only a small subset of an individual's actual peer group and therefore these studies are likely to have underestimated the total magnitude of peer effects due to measurement error in the peer group. Lyle (2007) and Carrell et al. (2009) explored peer effects in educational outcomes using college students who were randomly assigned to a peer group in which they did not only live in adjacent dorm rooms but also had to spend most of their study and leisure time together. Lyle (2007) found that there are positive peer effects in first year's achievements but that occurrences that are common to the group, the "common shocks", account for half or more of the estimated peer effect. Carrell et al. (2009) suggested that their approach identifies and well measures the true peer group and found strong and robust academic peer effects which were much larger in magnitude than in the previous literature. Carrell et al. (2009) showed empirical evidence that roommates and dorm floors capture only a limited proportion of the total peer influence finding only moderate evidence of peer influence at the roommate level, as previously found by Sacerdote (2001) and Zimmerman (2003), and also that geographic proximity of students in dorm halls alone, as in Foster (2003), does not generate measurable peer effects.

Differently from previous studies which explored peer effects in Higher Education using

samples from the USA, Brunello et al. (2010) exploited the random assignment of first year students in a middle-sized public university in Italy. The authors found that roommate peer effects vary with the field of study and particularly that peer effects for freshmen enrolled in the pure sciences are positive and significantly larger than for freshmen enrolled in the humanities and social sciences.

As mentioned above, there is a limited literature which used the random assignment technique to explore peer effects in school students. An exception is the paper of Sanbonmatsu et al. (2005) who used random assignment on school students, analysing the consequences of randomly changing the residential neighbourhood of families residing in high-poverty public housing. Using the experimental design of a housing mobility project, named Moving To Opportunity (MTO), allowed the authors to address the selection problem using a randomised design and therefore isolate the impact of residential neighbourhood characteristics on educational outcomes which were measured from achievements on the Woodcock Johnson tests of cognitive abilities. The authors found no statistically significant effects on test scores for any age group of individuals who were assessed four to seven years after randomisation suggesting that achievement-related benefits from improved neighbourhoods are small.

Several other studies investigated peer effects in secondary education by exploiting natural experiments. Particularly, Boozer and Cacciola (2001), Graham (2008) and Sojourner (2013) used the US experimental study of class-size reduction (Project STAR) to examine peer effects caused by differing class sizes. Boozer and Cacciola (2001) and Graham (2008) reported significantly positive peer effects within classrooms while Sojourner (2013) found moderate positive peer effects. Sojourner (2013) also explored heterogeneous peer effects and found evidence suggesting that lower-achieving students benefit more than higher-achieving students from increases in peer ability[13]. Further, Angrist and Lang (2004) used Boston's METCO program, which sends black students out of Boston's public schools into the more affluent suburbs, as an exogenous source of variation in peer ability by analysing exogenous changes in classroom compositions and indicated only limited evidence of statistically significant results.

Fixed Effects

The random allocation of students became very popular for peer effects in Higher Education

[13]It is worth mentioning that the true randomness of this experiment has been questioned (Hanushek, 1999).

but it has not been adopted in school level achievements because of the difficulty in finding evidence of random allocation within schools. For school age students several studies have adopted the fixed effects technique applying school, teacher, student, school-year and school-by-grade fixed effects. As explained by Sacerdote (2011), the basic concept of these studies is that by applying fixed effects in the estimation the selection effects are removed and this allows the researcher to identify peer effects from idiosyncratic variation in peer ability.

One of the most popular studies is the one from Hanushek et al. (2003) who used data from Texas and controlled for fixed student and school-by-grade effects to show that peers' achievements have a positive effect on individual grades and that this remained constant across quartiles of the grade distribution. Several other papers adopted this approach. Betts and Zau (2004), particularly, examined the impact of classroom and grade level peer achievement on individual elementary students' rate of achievement using a detailed panel data-set from San Diego Unified School District. The authors employed student level fixed effects to control for positive tracking of students into classrooms and found positive and significant peer effects. In reference to these studies, though, Atkinson et al. (2008) commented that there are still concerns that there could be non-random allocation of pupils to classes within schools and also non-random allocation of teaching resources including teacher quality to classes.

Burke and Sass (2013), went one step further and exploited non-linear peer effects by separating students by their own ability level and by applying school as well as teacher fixed effects. The authors found sizeable and significant peer effects and concluded that students with low initial achievement levels benefit less from an increase in the average ability of their peer group while middling and high initial achievement levels do best when placed with high ability students. The findings of their study suggested that classroom assignment policies involving some degree of tracking by ability should be preferred to policies in which all classrooms contain a broad mix of students. Similarly, Hoxby and Weingarth (2006) took advantage of transitory fluctuations in school composition which caused students to experience new peers in the classroom. Conditioning on students' fixed effects the authors estimated moderate to large peer effects and also mentioned that peers' background characteristics such as race, ethnicity and income have only slight effects once peers' achievement is properly accounted for. Further, Lin (2010) attempted to identify both endogenous and exogenous effects using spatial auto-regressive models with group

fixed effects. Applying the model to data sets from the National Longitudinal Study of Adolescent Health the author found strong evidence for both endogenous and exogenous effects in students' achievements.

In contrast with the above studies, the study of Vigdor and Nechyba (2007) which used data covering all public school students in North Carolina and employed school fixed effects, identified positive and significant peer effects. Nevertheless, the estimates under alternative specifications that exploited changes in school composition, particularly by employing teacher fixed effects, became negative and statistically significant. As discussed by Sacerdote (2011), the findings of that study call into question the methodology of using school fixed effects to identify peer coefficients.

Several additional studies estimated peer effects using samples outside of the US[14]. Lavy et al. (2012) found positive peer effects and further explored non-linearities which showed that the proportion of high achieving peers in class has no effect on the academic performance of most regular students but it does affect positively the outcomes of the brightest among the regular students. Also, they found that the proportion of low achieving peers had a negative effect on the performance of regular students. Ammermueller and Pischke (2009) examined peer effects for fourth graders in six European countries and found positive and significant peer effects which were modestly large. A limitation of this study is that it does not directly measure the academic ability of students' peers but relies on socio-economic background characteristics as proxies for this. Further, McEwan (2003) using evidence from a sample of 8^{th} graders in Chile found large effects from peers' background characteristics as for example the classroom mean of mothers' education.

Kang (2007) investigated the existence and structure of academic interactions within classrooms using a unique quasi-randomisation that takes place in the allocation system of middle school students (age 13) in South Korea. Using this randomisation in student placement within classrooms the author found that the mean achievement of classroom peers was positively correlated with a student's performance and in addition, using quantile regression, found that weak students interacted more closely with other weak students than with strong students and therefore their learning was more greatly affected and delayed by the presence of worst-performing peers. In contrast, strong students were found to interact more closely with other strong students and hence their learning could be improved by

[14]The most relevant studies are those examining UK samples and for that reason these studies will be discussed in a separate section below.

the presence of best-performing peers. The study suggested that weak students are likely to benefit from ability mixing, while strong students from grouping. Similarly, Ding and Lehrer (2007) using data from China showed that high achievers benefited most from increases in peer quality.

Instrumental Variables

Several authors attempted to solve the identification problem by designing instruments for peer behaviour that are assumed to be exogenous. Particularly, a straightforward implementation of this strategy is to design instrumental variables that model peer group characteristics as a function of exogenous variables (Evans et al., 1992). The attempts of previous studies to solve the identification problem using the instrumental variable approach are particularly relevant to this study as the instrumental variable approach is the main econometric technique adopted. The biggest concern for researchers following this approach is about the validity of the instrumental variables because it is hard to guarantee that they are correlated with the peer variables but are uncorrelated with the structural errors (Lin, 2010).

An example is the study of Rivkin (2001) who used the county group or metropolitan area characteristics as instruments for school level data and analysed peer group effects on high school achievements. The findings of his study suggested that aggregation tends to move estimates further from their true values and that these findings should raise strong doubts about the benefits of aggregation as a method to reduce selection bias. Further, Goux and Maurin (2007) used the individual and neighbourhood dates of birth within the year as a determinant of French children's early performance at school which is plausibly exogenous to the quality of the neighbourhood in which they live. In such a context, the authors tested whether children's performance at school is affected by the distribution of dates of birth within the year of the other children living in the same neighbourhood and found positive and significant peer effects. Particularly the authors found that regardless of the individual's own date of birth, children living in a neighbourhood with a relatively high proportion of children born at the beginning of the year perform significantly better than children living in a neighbourhood with a relatively high proportion of children born at the end of the year.

Further, de Xavier Pinto (2010) used a semi-parametric methodology and instrumented peer quality with the way that the students are allocated in classrooms. Precisely, the

authors used dummy variables representing the four ways that the school allocates students to classrooms which are integration by age, segregation by age, integration by score and segregation by score. They found evidence that peer effects are positive for students in the last year of primary school in Brazil and that students' test scores increase with student quality for average and high quality students. Fertig (2003) instrumented the coefficient of variation of peers with variables measuring whether a school selects students upon entry by placement tests or by their record of academic performance and whether the schools are in the private or public sector. The author reported that for US students the peer group composition is a strong predictor of individual achievement and also indicated that the higher the heterogeneity of achievement in a student's school, the lower is the individual performance suggesting that educational output is maximised in schools with a more homogeneous composition of students regarding their achievement.

Other studies which implemented the instrumental variable approach include the study of Dills (2005) who estimated peer effects using the introduction of a magnet school, which selects high ability students, into a school district as an exogenous source of variation in peer quality. The author explained that the introduction of this cream-skimming generates exogenous variation in the quality of classmates remaining to those students in the regular schools and thus minimises selection bias. The findings of this study showed that the loss of high ability peers lowered the performance of low-scoring students remaining in regular schools but the available data are limited to the ability of the students and do not control for any family or individual characteristic and therefore the findings cannot be taken as complete conclusions. Moreover, Lefgren (2004) used data from Chicago public schools and used as instrumental variables the variation in class setting policies arguing that this decision is school level and unrelated to student characteristics, after including school fixed effects, but does alter the pupil composition of classes. The author found peer effects to be quite small but generally positive and significant.

Several studies used the instrumental variable approach applied on UK samples. Mendolia et al. (2018) instrumented peers' achievements using the peers' of peers achievements and Gibbons and Telhaj (2016) used teacher expectations of attainments and the effectiveness of secondary school peers' origin primary schools as an instrument for pupil attainment. Further, Atkinson et al. (2008) instrumented peers' achievements using the average of a pupil's classroom peers' lagged attainment scores using school and teacher fixed effects and Bradley and Taylor (2004) used the random demographic change across catchment areas

(pupils moving between schools). The findings of these studies, which are of particular interest for the needs of this research, firstly because they use a sample from the UK and secondly because they use the same methodological approach that is adopted to this study, are discussed in more detail below.

Peer effects beyond achievements

As mentioned above, the literature on peer effects in education has been mostly concerned on outcomes which focus on educational achievements, usually measured from performance on tests and exams. There are few studies which examined peer effects in education beyond achievements. The study of Jonsson and Mood (2008), which is one of the closest in the literature with the present research in terms of outcome, examined possible peer effects on post-compulsory educational choices of two Swedish cohorts who were in the final grade of the comprehensive school in 1998 and 1999. The authors used a logit model and accounted for selection using school fixed effects to find results which support the social contrast theory, that the tendency to make a high-aspiring choice at upper secondary school is less for those who go to schools with high-aspiring peers when controlling for own achievement. Of course, the study of Jonsson and Mood (2008) cannot be considered an antecedent research of the present study as the methodological approach as well as the country coverage and time period differ. Facchinello (2017) also investigated a sample of Swedish 6^{th} graders to estimate the effect of classmates' ability on a student's compulsory school choices and found that an increase in average class ability reduces the probability of taking advanced Maths course while the English course choice was not affected by peers' choice.

Bobonis and Finan (2009) identified neighbourhood peer effects on children's school enrolment decisions using experimental evidence from the Mexican PROGRESA program. Using exogenous variation in the school enrolment of program eligible children to identify peer effects on the schooling decisions of ineligible children residing in treatment communities, the authors found that peers have considerable influence on the enrolment decisions of program-ineligible children and that these effects are concentrated among children from poorer households. The authors suggested that their findings imply that policies aimed at encouraging enrolment can produce large social multiplier effects.

Other studies investigating peer effects in education with outcomes other than school achievements, include De Giorgi et al. (2009), Sacerdote (2001) and Lyle (2007) who

examined peer effects on the choice of college majors. Further, Rivkin (2001) used instrumental variables to examine peer effects on high school continuation and high school non-participation. Mendolia et al. (2018) estimated the effect of secondary school peers on the likelihood to go to the university and further on the likelihood to attend a Russell-Group (selective) university. Evans et al. (1992) and Gaviria and Raphael (2001) examined peer effects on school drop-out decisions. While the older study of Evans et al. (1992) found insignificant neighbourhood peer effects on school drop-out decisions, the most recent study of Gaviria and Raphael (2001) found strong evidence of peer group effects at the school level. Table 3.1 below, presents a summary of the main studies examining peer effects in education on outcomes beyond achievements outlining their country coverage, the outcome that was examined and their main findings.

Table 3.1: Studies examining peer effects in education beyond achievements

Study	Country	Outcome	Peer effect
Jonsson and Mood (2008)	Sweden	Educational choices	Negative
Facchinello (2017)	Sweden	School subjects	Negative
De Giorgi et al. (2009)	USA	College majors	Positive
Sacerdote (2001)	USA	College majors	Positive
Lyle (2007)	USA	College majors	Positive
Bobonis and Finan (2009)	USA	School enrolment	Positive
Rivkin (2001)	USA	High school continuation	Insignificant
Evans et al. (1992)	USA	School drop-out	Insignificant
Gaviria and Raphael (2001)	USA	School drop-out	Positive
Mendolia et al. (2018)	England	University attendance	Positive

Peer effects using evidence from the UK

It is evident from the studies analysed above that peer effects, and particularly school level peer effects, vary a lot across different countries as the educational systems are different across them. For example some countries track students into differing-ability schools, others keep their entire school system comprehensive, other countries give to students the

chance to select their school and others do not. The most relevant literature to the present study is that examining peer effects using English/ British samples. The existing literature using British data is mostly based on data collected from the National Pupil Database (NPD) and School Annual Census which are both administrative datasets containing very limited information on pupils' background characteristics. Also, the current literature mostly analyses the impact of peers on secondary school achievements.

The studies of Gibbons and Telhaj (2016) and Lavy et al. (2012) exploit the change in peers from primary to secondary school and both find no significant evidence of linear-in-means effects for secondary school students in the UK. Both of these studies use the full set of UK students at age 14 for several recent cohorts. They use KS2 test scores (age 11) to measure peer inputs and KS3 test scores (age 14) to measure outcomes. For non-linear outcomes though, Lavy et al. (2012) found significant and sizeable negative peer effects arising from bad peers at the very bottom of the ability distribution but little evidence that average peer quality and very good peers significantly affect pupils' academic achievements. In addition, the authors found significant peer effects when separating the sample by gender, identifying that girls significantly benefit from the presence of very academically bright peers while boys marginally lose out. Gibbons and Telhaj (2016) even after applying a number of different methods concluded that peer effects exist but make a relatively small contribution to the variation in academic progress at that age.

Further, Bradley and Taylor (2004) attempted to restrict their sample only to students who changed school during the last two years of compulsory schooling since students who changed school also changed their peer group. The idea, as described by the authors, is that simultaneity bias is reduced when limiting the sample exclusively on movers since movers had less than two years to influence their peers compared to non-movers. To further reduce the impact of simultaneity bias the authors used lagged test scores of peers at the origin and destination schools. The findings of Bradley and Taylor (2004) reflected positive peer effects which were generally stronger for low and middle ability pupils than for high ability pupils and also determined that peer effects do not differ by gender but are substantially stronger for non-white boys than for other groups. However, the findings of this study could be doubted since, as Mendolia et al. (2018) discussed, pupils who change school may be systematically different from those who do not change, especially when the reasons for the change can be related to school achievements.

There are a few studies which estimate peer effects using longitudinal cohort studies. Particularly, Robertson and Symons (2003) used the NCDS cohort (born in 1958) and examined peer effects in Maths and English test scores improvement from age 7 to age 11. The authors measured the influence of the peer group by the percentage of the child's classmates who have fathers in the top socio-economic groups when aged 7 and controlled for selection bias by instrumenting the peer group effect with the individual's region of birth. Under these specifications, the authors found strong evidence for the importance of peer groups. Atkinson et al. (2008) used a panel of English pupils to look at the effect of the introduction of teacher's performance related pay. Conditioning on school and teacher fixed effects applied on a subset of schools that is argued to have random allocation of pupils, the authors found significant and non-trivial peer effects within the classroom. The instrumental variable methodology adopted by the authors as a robustness check of the random setting did not show significantly different results from the OLS estimations.

Finally, the recent studies of Mendolia et al. (2018) and Speckesser and Hedges (2017) are the closest methodologically to the present chapter. Mendolia et al. (2018), particularly, used the LSYPE dataset to examine school-level peer effects on GCSE and A level scores and on university attendance. They found that the average ability of peers has a moderate positive effect on GCSE and A level scores and that being in a school with a large proportion of low-quality peers can have significantly detrimental effects on individual achievements. Further, the authors found that peers' ability has stronger effect on students at the bottom of the grade distribution at age 16. Speckesser and Hedges (2017), adopting the same methodology as Mendolia et al. (2018), used English data from the NPD and Annual School Census to investigate peer effects on the decision to pursue an academic or vocational track after completing compulsory schooling. The authors found that individuals with peers achieving higher scores were less likely to choose a vocational track.

Table 3.2 summarises the main studies of peer effects in education using British data.

Table 3.2: Summary of key studies using British data

Study	Identification	Dataset	Outcome	Peer effect
Robertson and Symons (2003)	Instrumental Variables	NCDS	Maths and reading improvement	Positive
Bradley and Taylor (2004)	Random Assignment	NPD, Annual School Census	KS3, KS4 scores	Positive
Atkinson et al. (2008)	School and Teacher Fixed Effects, Instrumental Variables	NPD, Annual School Census	KS3, GCSE scores	Positive
Lavy et al. (2012)	Fixed Effects	NPD, Annual School Census	KS3 scores	Positive for heterogeneous outcomes
Gibbons and Telhaj (2016)	Instrumental Variables	NPD, Annual School Census	KS3 scores	Insignificant
Speckesser and Hedges (2017)	Instrumental Variables	NPD, Annual School Census	Academic Choice	Positive
Mendolia et al. (2018)	Instrumental Variables	LSYPE	GCSE, A level scores, University Attendance	Positive

3.2.3 Contribution to the literature

The difficulty in identifying the causal peer effect is reflected in the results of the numerous studies which are mixed; finding strong, weak or non-existent peer effects even when the finest econometric techniques are used. Some studies which used instrumental variables have been criticised as being ad-hoc in nature failing to identify variation that is more credibly exogenous and not based on a strong theory (Moffitt et al., 2001).

There are three main contributions of this chapter to the existing literature. First, the chapter adopts a novel 'peers-of-peers' identification strategy to overcome the possible selection encountered in the analysis of peer groups (Mendolia et al., 2018). Information on the primary school peers of an individual's secondary school peers who attended a different primary and secondary school from the individual of interest is used to instrument the peer effects. The idea is that these peers-of-peers could not have directly affected the individual's aspirations since they have never been in the same school with the individual. Second, the majority of recent studies on peer effects in education have been almost exclusively concerned with educational achievements. What makes this paper distinct from previous studies is that it is focused on the effects of peers on individuals' *aspirations* and *intentions*, rather than on their achievements. These are crucially important in understanding the mechanisms underlying the education production function. Third, the paper estimates peer effects in two different ways: through peers' average achievements and, for the first time, through peers' own aspirations.

In addition to these contributions, the paper adds to the existing literature on peer effects

in the UK by providing evidence from LSYPE. Unlike the majority of previous studies for the UK which used the limited information that is available in the National Pupil Database (NPD) and Annual School Census, the LSYPE dataset contains very rich information about the individuals, their family background and composition, and their schools.

3.3 Data

3.3.1 Description of the Dataset

The data are drawn from the LSYPE which forms the most recent cohort analysed in the previous chapter. A detailed description of the dataset has been provided in Subsection 2.3.1.

The analytical sample used in this chapter comprises of 7,938 pupils who have at least one secondary school peer and at least one peer-of-peer, as well as having non-missing information on educational aspirations reported at age 14 and age 16 and other information including demographics, family background and composition, and primary and secondary school characteristics including test scores at KS2 and KS3. The estimating sample did not differ significantly from the full sample in terms of background characteristics. Approximately 91% of the sample attended government comprehensive schools while 9% attended voluntary-aided or controlled schools, usually schools with a religious denomination. The students in the sample attended 533 different secondary schools, and had attended 3,445 different primary schools.

3.3.2 Peer Groups

The students in English secondary schools are grouped with different peers for different subjects and consequently they interact with most other students in their year group attending the same school rather than having a unique group of classmates as when in primary school. Therefore, the secondary school peer group of each individual includes all other students in that year attending the same school. While the school year group is a highly relevant peer group in the English context it is acknowledged that peer effects might be stronger from interactions with close friends than from the overall school group but unfortunately the available data do not allow us to identify friendships between the LSYPE individuals.

Under this setting, school fixed effects cannot be included in the estimated models as the

peer group would be part of the school fixed effect. In an attempt to eliminate this bias I control for class size and region but unobserved confounding factors are still likely to exist and therefore the estimated "peer effect" could pick up school effects rather than peer effects per se. For that reason, in order to identify the causal peer effect the chapter relies on using the peers-of-peers as an instrument for the secondary school peers.

Each LSYPE individual has a secondary school peer group which varies from 1 to 35 students observed in LSYPE. The average observed secondary school peer group of each LSYPE individual consists of 15 students. Their peers originate from between 2 to 23 different primary schools. The majority of secondary schools included students from between 7 to 14 different primary schools. Peers-of-peers are defined as students who went to the same primary school as individual i's secondary school peers, but a different primary school, and secondary school, to individual i. Table 3.1 below shows that over 77% of the individuals have a peers-of-peers group of 4 or more pupils who are observed in LSYPE.

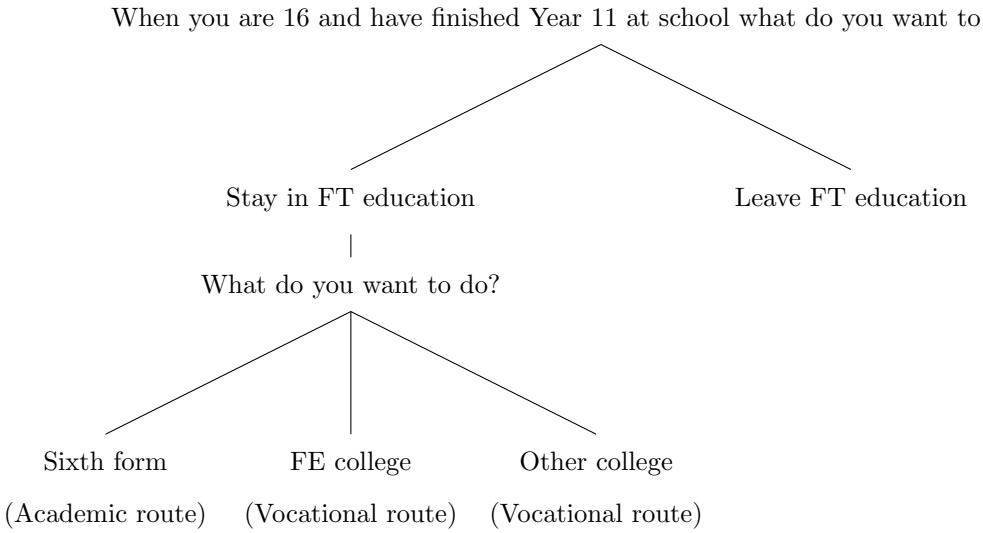
Table 3.1: Peers-of-peers Group Size

Number of peers-of-peers	% of LSYPE individuals
1 peer of peer	11
2-3 peers of peers	12
4-5 peers of peers	10
6-8 peers of peers	11
9-11 peers of peers	12
12-15 peers of peers	11
15+ peers of peers	33
Total	100

3.3.3 Variables

Dependent Variables

To study the influence of secondary school peers on individuals' educational aspirations, the future intentions of the LSYPE sample have been examined at two different time points. This enables us to observe both their preliminary and their later aspirations, and how these changed over the last two years of compulsory schooling. First, the individuals' *preliminary aspirations* were recorded when they were attending Year 9 of compulsory schooling at age 14. Then, their *later aspirations* were recorded when they were in Year 11 -the final year of compulsory schooling- at age 16. The derivation procedure of the dependent variables is illustrated in the tree diagram below:



Educational aspirations are measured in two ways. First, individuals are asked about their intentions to stay in full-time (FT) education or to leave education after completing compulsory education at age 16. Second, for those who are intending to stay in FT education, they are asked about their plans, choosing between attending a school 'sixth form' or a college of Further Education (FE). In England, whilst you can still study for academic qualifications at a FE college, the courses offered there tend to be more vocationally orientated as compared to school sixth forms which offer a range of more academically focused subjects[15]. Consequently, the individuals who mentioned a sixth form have been considered as aspiring to an academic route while those who mentioned a FE college have been considered as aspiring to a vocational route. Table 3.2 below provides the descriptive statistics for the four dependent variables.

Overall, 91% of the individuals were aspiring to stay in FT education both at age 14 and age 16. Figure 3.1 shows the cross-tabulation of the preliminary and later aspirations for staying or leaving FT education. Interestingly, around 60% of the individuals who intended to leave FT education at age 14 responded that they intended to stay when asked again at age 16. For individuals who indicated that they intended to stay in FT education, 74% were aspiring to an academic route at age 14 while at age 16 this proportion decreased slightly

[15]While it is true that individuals attending a college of FE, rather than a sixth form, tend to be more likely to follow a vocational route as they are exposed to a greater extent to a variety of vocational subjects, it is still possible for them to follow an academic route or to continue with academic education in a university after achieving vocational qualifications. Although ate age 13-14, when the preliminary aspirations were measured, the individuals did not specify what qualifications they were aspiring, in their later aspirations, at age 15-16, the individuals were asked to specifically determine what qualifications they aspired to do and whether they wanted to study for these qualifications in order attend a university. Using this information an additional dependent variable has been examined, Aspiration 5, which considers that the individuals aspired to the academic route if they mentioned A levels and/or university attendance and a vocational route if they mentioned other qualifications. This variable is constructed in an identical way to the aspirations variable used in the previous chapter and an exact illustration of the derivation procedure can be found in Appendix A.1. The outcome from this analysis is presented and discussed in Appendix B.5.

Table 3.2: Descriptive Statistics of Dependent Variables

Educational Aspirations	Mean	Std. Dev.
Preliminary Aspirations (age 14)		
A1: Stay in FT education	0.91	0.29
A2: Academic route | *conditional on aspiring to stay*	0.74	0.44
Later Aspirations (age 16)		
A3: Stay in FT education	0.91	0.29
A4: Academic route | *conditional on aspiring to stay*	0.70	0.46

All variables are dichotomous. The alternative categories for aspirations A1 and A3 are: Leave FT education; for aspirations A2 and A4: Vocational route

to 70%. It is evident from these raw data that the academic route is much more strongly preferred to the vocational route. Figure 3.2 shows that almost half of the individuals who stated that they aspired to a vocational route at age 14 changed to intending to follow an academic pathway at age 16. In contrast, only one fifth of the individuals who were aspiring to an academic route at age 14 changed to a vocational aspiration by age 16.

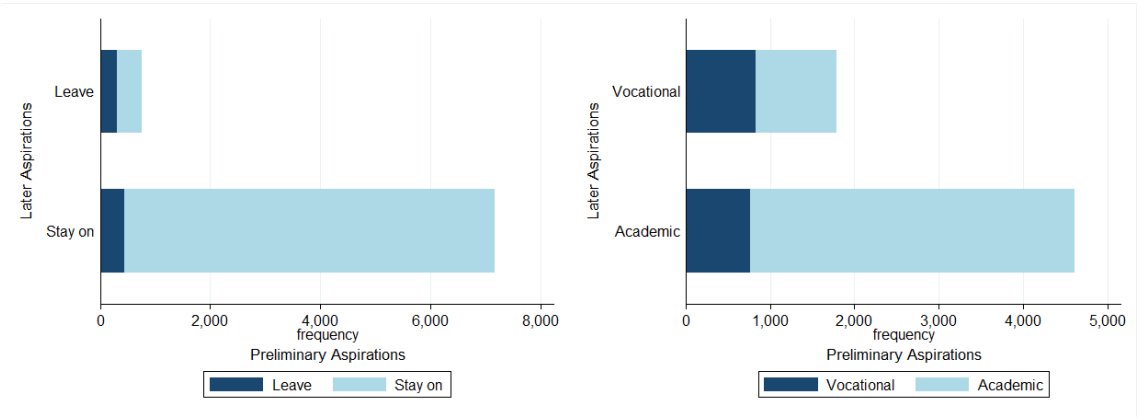


Figure 3.1: Leave or Stay in FT education **Figure 3.2: Academic or Vocational route**

Key Variables

The primary aim of this chapter is to analyse the effects of peers on individuals' educational aspirations. The influence of the peer group is measured in two different ways: through peers' average ability as reflected in their educational achievements and through peers' own aspirations. The possible selection into peer groups is controlled for using the average ability of the peers-of-peers as an instrument for the average ability of peers. Peers' average aspirations are instrumented using both the average ability and the proportion of peers of peers from the highest socio-economic group. Table 3.3 below reports the summary statistics of the key variables determining the peer effects.

Table 3.3: Descriptive Statistics of Key Variables

Key Variable	Mean	Std. Dev.
Secondary School Peers		
Average KS3 scores (standardised)	0	1
Preliminary Aspirations (age 14)		
A1: Stay in FT education	0.89	0.10
A2: Academic route	0.73	0.21
Later Aspirations (age 16)		
A3: Stay in FT education	0.90	0.10
A4: Academic route	0.69	0.22
Peers-of-Peers		
Average KS2 scores (standardised)	0	1
Proportion from highest socio-economic groups	0.19	0.14

Peers' ability

Peers' ability is measured by average secondary school peers' achievements in KS3 exams (Maths, English, Science) in standardised form. To avoid possible selection in peer groups, peers' ability is instrumented by the average ability of the primary school peers-of-peers, measured by average achievements in KS2 exams (Maths, English, Science) in standardised form.

Peers' educational aspirations

Peers' educational aspirations are measured by the average secondary school peers' aspirations, ranking from 0 to 1, with values closer to 1 indicating that more peers are aspiring to stay in FT education (for aspirations A1 and A3) or to follow an Academic route (for aspirations A2 and A4). Figures 3.3-3.6 present the distribution of peers' aspirations. These figures show large variation in aspirations across schools, particularly in the aspirations to follow an academic or vocational route. Since there are no observations on the aspirations of primary school students, the possible selection in peer groups has been handled by instrumenting peers' aspirations with peers-of-peers' ability and socio-economic background.

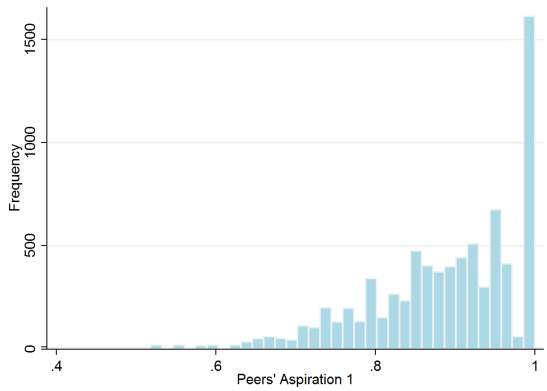


Figure 3.3: Peers' A1: Stay in FE ed

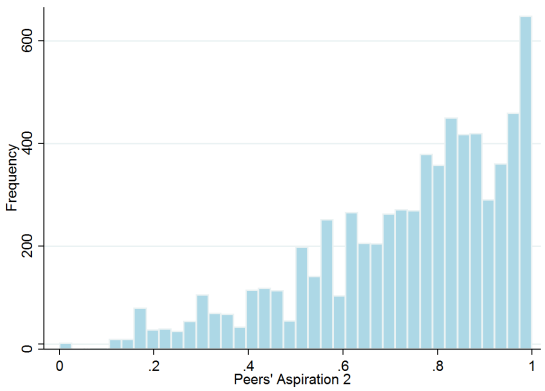


Figure 3.4: Peers' A2: Academic route

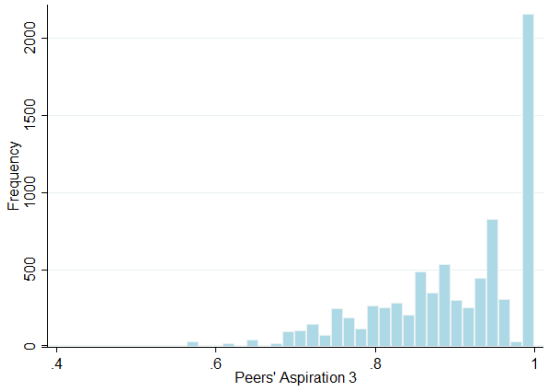


Figure 3.5: Peers' A3: Stay in FT ed

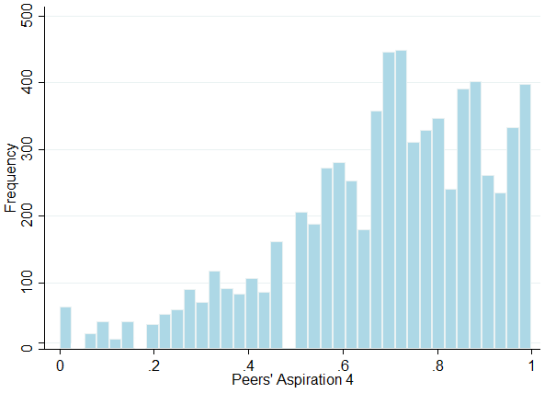


Figure 3.6: Peers' A4: Academic route

Control Variables

The LSYPE dataset is very rich in background information about the individual and their family and household as well as the characteristics of the secondary school attended. This information is included in the regression analyses to control for other conflating factors which may impact upon an individual's educational aspirations in order to identify the impact of peers' (achievements, socio-economic background and aspirations) on the individual's educational aspirations.

Three variants of the model have been estimated, progressively adding more covariates. The first specification controls for individual demographics, including gender, ethnicity, ability (as reflected in their KS2 scores), and the education and employment status of both parents. The second specification adds household composition characteristics including the age of both parents, parents' marital status and the number of siblings. The third specification adds school's class size and geographic region. The descriptive statistics for the control variables are provided in Appendix B.1.

3.4 Identification Strategy

To investigate peer effects on educational aspirations, the chapter estimates a linear-in-means model for the i^{th} individual who attended secondary school s and primary school p:

$$Y_{i,s,p} = \beta_0 + \beta\overline{A}_s + \gamma X_i' + \epsilon_i$$

where $Y_{i,s,p}$ is the aspiration of individual i who attended secondary school s and primary school p, $\overline{A}_s$ is the peer effect for secondary school s excluding the individual i and X_i' is a vector of individual, household, family and school characteristics for individual i.

The parameter of interest is β which captures the relationship between secondary school peers and individual i's aspirations. As the estimates of β are likely to be biased due to selection in peer groups, the identification strategy relies on instrumenting these peers with the peers-of-peers. In particular, the chapter uses the primary school peers of the individual's secondary school peers who attended a different primary school from the individual as an instrument for their secondary school peers. Peers' achievements in KS3 exams have been instrumented using peers-of-peers' achievements in KS2 exams. Peers' aspirations have been instrumented using peers-of-peers achievements in KS2 exams and their SEG.

The first stage equation is:

$$\overline{A}_s = \delta_0 + \delta\overline{K}_{q,r} + \pi X_i' + \nu_i$$

where the average peer effect $\overline{A}_s$ depends on the peers-of-peers who attended primary school q and secondary school r, where $r \neq s$ and $q \neq p$. The defining point is that the peers-of-peers have not been in either the same primary school or secondary school as the individual of interest. The underlying assumption behind this identification strategy is that the peers-of-peers can not have affected the individual's aspirations directly, but rather only through their effect on the individual's current secondary school peers[16].

[16]There is a potential threat to identification with the current construction of the instrumental variables. The instrumental variables identify for each individual their secondary school peers who went to different primary schools and then uses the average KS2 scores of their primary school peers and the proportion of their primary school peers from the highest socio-economic group. The potential bias arises when an individual has secondary school peers who where in the same primary school (but different from his or her primary school). For example, if an individual's secondary school peer A was in primary school with his or her secondary school peer B and this primary school is not the same as his or her primary school,

Any bias driven by selection into secondary schools on the basis of unobservables (for example, parents choosing the school for their children), is mitigated by the nature of the primary to secondary school transition in England. Most secondary schools participating in LSYPE have more than eight primary school feeders and therefore the peers-of-peers are likely to have come from areas with different socio-economic characteristics. Further, Gibbons et al. (2013) showed that neighbourhood composition in England has a very limited effect on individual achievements once own family background is accounted for. The very rich set of control variables afforded by the LSYPE means that the chapter can successfully account for such factors.

3.5 Empirical Results

3.5.1 Main findings and robustness checks

The main results are reported in Table 3.1, which shows the coefficient of interest (β) on the peer effect variable in various specifications. The equations for each aspiration are estimated using both the Linear Probability Model (LPM) and IV-Two Stage Least Squares (IV-2SLS) regressions[17], using instruments for the peer effects as described above. All of the estimated models use robust standard errors clustered at the secondary school level. For each method, there are three specifications that condition on increasingly adding more covariates, as described in Section 3.4 above, and summarised in the footnotes to the table. The first stage results of the IV-2SLS estimations in each specification together with their F-statistics, verifying the validity of the instruments, are presented in Appendix B.2. To summarise the key message to come out of Table 3.1, there is strong evidence for positive peer effects influencing the decision to choose an academic rather than vocational route in both preliminary (at age 14) and later aspirations (at age 16) amongst those who have decided to continue with their education (aspirations A2 and A4). This result is independent of how the peer effects are measured, the additional factors controlled for in the regression specification, and whether or not the peer effects are treated as endogenous. The evidence is clear that studying with peers with a higher level of prior achievement

then peer B will be included in the primary school peer group of peer A and peer A will be included in the primary school peer group of peer B. This is a limitation of the current construction of the instrumental variable. It is not a cause of major bias in the estimation given that secondary school peers KS2 scores are pre-determined (meaning that they were taken before the secondary school peers met the individual) and also that there are not many such cases in the data since most secondary schools participating in LSYPE have more than 8 primary school feeders.

[17]The models have been also estimated using Probit and IV-Probit regressions given the aspirations are all binary (1.0) variables, and the findings remain unchanged.

or peers whose own aspirations are to follow an academic route, are all associated with a higher likelihood that an individual will aspire to academic post-compulsory education themselves.

Considering the results in more detail, focusing first on the IV results for the academic vs. vocational aspirations, then at age 14 (column 4), a 1 standard deviation (sd) increase in an individual's peers' KS3 scores is associated with an 3-5 percentage point (pp) increase in the likelihood that they aspire to an academic rather than a vocational post-compulsory education pathway. The range of effects is determined by the number of other factors controlled for, with the most extensive range of controls being associated with the estimate at the bottom of the range. In addition, each 10 pp increase in the proportion of an individual's peers who aspire to the academic route is associated with a 5 to 6 pp increase in the likelihood that the individual has similar aspirations themselves. At age 16 (column 8), the estimated peer effects are very similar for the prior attainment measure, although are slightly smaller in magnitude for the socio-economic background and peers aspirations categories.

Comparing the LPM to the IV-2SLS results, the IV effects are slightly smaller in magnitude in most cases while there are some cases where the IV estimates are actually larger than the LPM effects. Collectively, however, the results suggest that the selection bias is small.

Turning to the decision whether to continue in education after the age of 16, the evidence that individuals' intentions are influenced by their peers is much weaker. At age 14 (Aspiration 1), the LPM results (column 1) reveal positive coefficients on the peers socio-economic background and aspirations variables, but these are small in magnitude, and are statistically insignificant for the peer achievement measure. Furthermore, these coefficients all become even smaller, and statistically insignificant in every case (column 2) when the IV methodology is applied, suggesting that any positive correlation that is observed in column 1 is in fact more likely due to self-selection into peer groups, than to true causal effects of the peers. Two years later, when pupils are aged 16 and closer to making their decision whether to continue in FT education or not (Aspiration 3), there is evidence that positive peer effects are present even after instrumenting (column 6) for one of the peer effects measures, although these effects are small in magnitude. As Table 3.2 showed earlier, around 90% of pupils intend to continue in FT education, and so for most this would appear to be almost an automatic choice, regardless of their peers. It may be

that at the margin, some pupils are influenced by the aspirations of those around them, but for most it is not a marginal decision, and so such effects are small in the aggregate results presented here.

The full regression results including the other independent variables are reported in Appendix B.3. As expected, family socio-economic background (particularly parents' education) and individual's own achievements in KS2 exams are strong determinants of educational aspirations. In addition, girls, pupils from Black/Asian backgrounds, and those who live in London are found to be more academically ambitious, aspiring to stay in education and to follow an academic route. Further, individuals with fewer siblings and those whose parents are married are also found to have stronger intentions to stay in FT education, and to follow an academic route. Interestingly, larger class size is found to negatively affect an individual's aspirations to stay in education and to follow an academic route.

As reported in Appendix B.4, the robustness of these findings has been tested using three sensitivity analyses on the most detailed model (Specification 3). First, the model has been re-estimated excluding observations from very small secondary schools (with fewer than 700 students). As explained by Mendolia et al. (2018), larger secondary schools typically draw their intake from a greater number of primary schools and this is likely to lessen the problem associated with socio-economic sorting in primary schools. Secondly, the main results have been re-estimated limiting the sample to schools who have at least 10 LSYPE individuals. Both of these sensitivity analyses corroborate the main findings in Table 3.1. The final sensitivity analysis involved sorting the sample by the number of peers-of-peers they have and then re-estimating the main results using different sub-samples, progressively excluding individuals with smaller numbers of peers-of-peers. The aim of this analysis is to show the approximate stability of the findings as individuals with few peers-of-peers are progressively excluded and also to identify whether the results could be affected by the fact that the chapter only uses a small sample (LSYPE participants) of the total cohort of individuals who finished school in 2006. Interestingly, the outcome of this analysis shows that peer effects become stronger when the sample is limited to individuals who have many peers-of-peers.

3.5.2 Heterogeneous peer effects and peer effects on the probability of pupils changing their aspirations between Year 9 and Year 11

Some of the findings might be driven by the fact that the effect of peers on individual aspirations is heterogeneous, such that peers' influence might be more important for particular groups of pupils. For example, some pupils might be more heavily influenced by the existence of higher achieving peers while others not. For this reason potential heterogeneity of peer interactions is examined, first, by estimating peer effects by gender and second, by examining peer effects for pupils who received different types of information, advice and guidance (IAG) concerning their educational plans and future aspirations. Finally, the chapter examines peer influence on whether pupils change their aspirations between age 14 and age 16. For ease of exposition, only results from Specification 3 are presented for all of these analyses.

Tables 3.2 and 3.3 present results separately for girls and boys respectively. For preliminary aspirations to follow an academic or vocational route in post-compulsory education (aspiration A2), the peer effects are quite similar for girls and boys though they appear to be slightly larger in magnitude for boys. The peer effects on later aspirations to follow an academic or vocational route in post-compulsory education (aspiration A4) become smaller and less precise for girls while remaining positive and significant for boys. Interestingly, for individuals' later aspirations to continue into post-compulsory education (aspiration A3), all peer effects when estimated by IV, are small and statistically insignificant for girls while positive and significant for boys. Considering the gender analysis all together, the findings suggest that peer influences on educational aspirations appear to be stronger on boys than on girls.

Tables 3.4-3.7 present the analysis of peer effects for the pupils who received educational IAG compared with those pupils who did not receive any advice. Specifically, the pupils are separated into four groups. Those who received no advice, those who received advice from a Connexions Personal adviser or someone else at Connexions, those who received advice from a careers adviser/ teacher or other teacher at school and those who received both Connexions and teacher advice[18]. The analysis of peer effects on the preliminary

[18]Connexions was a UK governmental information, advice, guidance and support service for young people aged 13-19, created in 2000 following the Learning and Skills Act. There were several Connexions Centres in each county which offered support and advice on topics including education, housing, health, relationships, drugs, and finance. Connexions is no longer a coherent National Service following the announcement of changes to the delivery of careers in England by the Coalition government. A 2010 research report by the National Youth Agency and the Local Government Association noted that some young people were

aspirations of pupils who received advice compared to those who did not, has been examined based on information concerning whether they received advice by the end of Year 9 of compulsory schooling. Similarly, peer effects on the later aspirations of pupils who received advice compared to those who did not, are examined based on whether the pupils received advice by Year 11 which is their final compulsory school year. As shown by the number of observations for each sub-group of pupils, presented in Tables 3.4-3.7, about 47% of the pupils had not received any IAG by Year 9, when their preliminary aspirations were recorded, while by Year 11 only 5% of the pupils had received no advice. The majority of pupils in Year 11 (62%) received both Connexions and school advice.

The outcome of this analysis is interesting. Peer effects on pupils' preliminary aspirations to follow an academic or a vocational route (Aspiration 2) are statistically insignificant for most pupils who received IAG, while they remain positive and statistically significant for the pupils who received no advice. Peer effects on the pupils' later aspirations (Aspiration 4) are still positive and significant for those who received no advice and for the students who received only school advice while becoming statistically insignificant for the pupils who received only Connexions advice and those who received both school and Connexions advice[19]. For pupils' early and later aspirations of whether to stay in education or not (Aspiration 1 and Aspiration 3), peer effects are statistically insignificant for all groups of pupils[20].

The final analysis presented in Table 3.8 investigates peer influences on whether pupils change their aspirations between Year 9 and Year 11 of compulsory schooling (between age 14 and age 16). The results show that having higher ability peers is associated with a lower likelihood of an individual changing aspirations between Year 9 and Year 11, especially regarding the choice between academic and vocational pathways, conditional on the intention to stay in FT education (Aspiration 2). Thus, not only do peers influence individuals' intentions for undertaking academic rather than vocational post-compulsory education, but they also make their aspirations less likely to vary over time.

unclear about the role and function of Connexions, although those who had interacted with the service were generally positive about it (Hibbert, 2010). Connexions was external to schools and therefore a visit to a Connexions advisor was an active choice in contrast to careers/ teacher advice at school which is more random, from the point of view of the pupils, depending on the schools' policy. Therefore, the chapter considers separately the pupils who received only Connexions advice from those who received only school advice and those who received both forms of advice.

[19]It is acknowledged that in this case where the influence of peers on aspirations becomes insignificant, it could still be the case that the individuals were inspired by their peers to obtain the Connexions advice in the first place.

[20]There is one exception on the later aspirations of pupils who received both school and Connexions advice where I still find a positive peer influence for one of the peer effects measures.

3.6 Conclusion

This paper has established the existence of strong causal effects of peers on pupils' aspirations about their post-compulsory education, in particular whether to follow an academic rather than a vocational route. The results show that pupils who attend school with higher achieving peers and with peers who are more likely to aspire to an academic post-compulsory education, are more likely to aspire to follow an academic route themselves. The causality is established by using an IV procedure to account for the endogenous nature of peer group formation and selection. The chapter takes advantage of the fact that the linked administrative data identify both the primary school and the secondary school attended by each LSYPE respondent. This enables the chapter to utilise the 'peers-of-peers', i.e. pupils who attended the same primary school as an individual's peers, but a different primary and secondary school to the individual themselves.

The results are consistent with many studies that have found evidence for peer effects. In particular, they are consistent with the findings of Mendolia et al. (2018) and Speckesser and Hedges (2017), who also consider secondary school pupils in England. The chapter adds to this literature by focusing on aspirations, showing for the first time that peers are important in forming individuals' post-compulsory educational aspirations. Given the importance of aspirations for eventual outcomes, the results can therefore help to explain the relationships between peers' outcomes that were observed in these earlier studies. In particular, the results presented in this paper show that peers can influence the aspirations of girls and, particularly, boys to follow an academic pathway post-16, conditional on having decided to continue with their education. Further, interestingly enough, the findings indicate that peers' aspirations have a greater impact on individuals' aspirations that peers' achievement.

The results have implications for allocations of pupils across schools. Even in a mostly comprehensive education system as in England, with no selection by schools on ability, there are still large differences in pupil intakes across schools, in terms of prior ability, as shown in Section 3.3 above. This is mostly associated with clustering of families by background, with better-off families able to pay higher house prices closer to high-performing schools (Gibbons and Machin, 2003).

The analysis does not pass judgment as to whether the peer effects on choice of route benefit the individual in question or not. While in some cases individuals may be inspired to

undertake a route that turns out to be beneficial but which they might not otherwise have chosen, in other cases, they may be influenced by their peers to take a less advantageous or appropriate route. For example, some individuals could follow their academic-orientated peers when a vocational course may have been more suitable for them, while other individuals in a vocationally-dominant peer group may be more suited to academic study themselves. In such cases where an individual could be influenced into making the 'wrong' choice for their own personal circumstances by simply following their peers, the analysis of peer effects on pupils who received educational IAG suggests that such advice could play an important role given that it is shown to weaken the influence of peers. Given the difficulty of identifying a priori those pupils who would make the 'wrong' choice if following their peers, this suggests the importance of providing educational advice and guidance to all pupils as an alternative source of information to guide choices. Further, even if it is particularly complex to draw clear policy implications related to students ability mixing, ideally, when considering educational aspirations the influence of peers ability should be zero so that young individuals can make the best choices in relation to their own ability. As a consequence, secondary school composition at transition from KS2 to KS3 should be more carefully looked at.

Table 3.1: LPM and IV-2SLS estimates of peer effects on educational aspirations

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT ed		Acad/Voc route		Stay/Leave FT ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV 2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer effect: Average peers' achievements in KS3 exams								
Specification 1	0.004	0.005	0.055***	0.050***	0.012***	0.011*	0.036***	0.040***
	(0.004)	(0.006)	(0.008)	(0.014)	(0.003)	(0.006)	(0.008)	(0.014)
Specification 2	0.003	0.004	0.052***	0.045***	0.009***	0.007	0.033***	0.036***
	(0.004)	(0.006)	(0.008)	(0.014)	(0.003)	(0.006)	(0.008)	(0.014)
Specification 3	0.003	0.004	0.044***	0.035***	0.009**	0.006	0.025***	0.029**
	(0.004)	(0.007)	(0.008)	(0.014)	(0.003)	(0.006)	(0.007)	(0.013)
Peer effect: Average peers' aspirations								
Specification 1	0.220***	0.159	0.747***	0.641***	0.249***	0.345***	0.633***	0.400***
	(0.043)	(0.118)	(0.021)	(0.081)	(0.044)	(0.117)	(0.027)	(0.094)
Specification 2	0.220***	0.128	0.741***	0.592***	0.237***	0.264**	0.626***	0.400***
	(0.043)	(0.124)	(0.021)	(0.085)	(0.044)	(0.123)	(0.027)	(0.094)
Specification 3	0.212***	0.131	0.707***	0.551***	0.213***	0.232*	0.592***	0.340***
	(0.045)	(0.130)	(0.023)	(0.095)	(0.048)	(0.137)	(0.029)	(0.108)
Observations	7938	7938	6950	6950	7938	7938	7053	7053

Robust standard errors clustered at the secondary school level are reported in in parentheses.

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

Specification 1: gender, ethnicity, KS2 scores, parents' employment status, parents' education.

Specification 2: Specification 1 plus parents' age, number of siblings, parents' marital status.

Specification 3: Specification 2 plus geographic region and class size

Table 3.2: LPM and IV-2SLS estimates of peer effects on educational aspirations of girls

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT ed		Acad/Voc route		Stay/Leave FT ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.007	0.008	0.044***	0.032*	0.009	-0.005	0.027***	0.019
	(0.005)	(0.010)	(0.010)	(0.017)	(0.005)	(0.011)	(0.009)	(0.016)
Peers' Average Aspirations	0.179***	0.201	0.744***	0.515***	0.250***	-0.025	0.568***	0.287*
	(0.068)	(0.183)	(0.035)	(0.157)	(0.065)	(0.249)	(0.041)	(0.161)
Observations	3895	3895	3270	3270	3895	3895	3309	3309

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Included but not shown: Specification 3 variables.

Table 3.3: LPM and IV-2SLS estimates effects of peer effects on educational aspirations of boys

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT ed		Acad/Voc route		Stay/Leave FT ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	-0.001	-0.003	0.044***	0.049**	0.007*	0.014*	0.022**	0.039**
	(0.005)	(0.008)	(0.009)	(0.018)	(0.004)	(0.008)	(0.010)	(0.019)
Peers' Average Aspirations	0.225***	-0.020	0.679***	0.597***	0.174**	0.375**	0.606***	0.382***
	(0.052)	(0.189)	(0.036)	(0.125)	(0.068)	(0.174)	(0.039)	(0.144)
Observations	4043	4043	3680	3680	4043	4043	3744	3744

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Included but not shown: Specification 3 variables.

Table 3.4: LPM and IV-2SLS estimates of peer effects on educational aspirations of pupils who received no advice

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT ed		Acad/Voc route		Stay/Leave FT ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	-0.000	0.006	0.042***	0.060***	0.040***	0.040	0.049**	0.068*
	(0.006)	(0.009)	(0.009)	(0.017)	(0.015)	(0.029)	(0.022)	(0.039)
Peers' Average Aspirations	0.222***	0.158	0.740***	0.725***	0.260	0.785	0.518***	0.627**
	(0.065)	(0.175)	(0.037)	(0.131)	(0.174)	(0.673)	(0.115)	(0.285)
Observations	3319	3319	2894	2894	424	424	377	377

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Included but not shown: Specification 3 variables.

Table 3.5: LPM and IV-2SLS estimates of peer effects on educational aspirations of pupils who received only Connexions advice

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT ed		Acad/Voc route		Stay/Leave FT ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	-0.004	-0.005	0.053***	0.030	0.014*	0.017	0.022*	0.024
	(0.007)	(0.012)	(0.013)	(0.025)	(0.008)	(0.016)	(0.012)	(0.022)
Peers' Average Aspirations	0.113	0.064	0.652***	0.268	0.551***	0.433	0.564***	0.260
	(0.077)	(0.337)	(0.053)	(0.234)	(0.139)	(0.375)	(0.061)	(0.228)
Observations	1949	1949	1713	1713	1610	1610	1360	1360

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Included but not shown: Specification 3 variables.

Table 3.6: LPM and IV-2SLS estimates of peer effects on educational aspirations of pupils who received only school advice

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT ed		Acad/Voc route		Stay/Leave FT ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.002	-0.006	0.044**	0.002	0.003	-0.006	0.023	0.061**
	(0.009)	(0.017)	(0.019)	(0.034)	(0.009)	(0.017)	(0.015)	(0.025)
Peers' Average Aspirations	0.201	-0.133	0.713***	0.085	-0.029	0.019	0.521***	0.599***
	(0.126)	(0.373)	(0.080)	(0.412)	(0.086)	(0.332)	(0.087)	(0.211)
Observations	817	817	744	744	938	938	856	856

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Included but not shown: Specification 3 variables.

Table 3.7: LPM and IV-2SLS estimates of peer effects on educational aspirations of pupils who received both Connexions and school advice

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT ed		Acad/Voc route		Stay/Leave FT ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.009	0.004	0.044***	0.015	0.007*	0.001	0.025***	0.022
	(0.009)	(0.018)	(0.016)	(0.031)	(0.004)	(0.008)	(0.009)	(0.016)
Peers' Average Aspirations	0.211**	0.064	0.674***	0.527*	0.130***	0.199	0.602***	0.274*
	(0.097)	(0.499)	(0.072)	(0.289)	(0.050)	(0.179)	(0.034)	(0.158)
Observations	1026	1026	925	925	4884	4884	4397	4397

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Included but not shown: Specification 3 variables.

Table 3.8: LPM and IV-2SLS estimates effects of peer effects on the probability of changing educational aspirations between Year 9 and Year 11 of school

	Changed Stay/Leave Aspiration		Changed Ac/Voc Aspiration	
	LPM	**IV-2SLS**	**LPM**	**IV-2SLS**
Peers' Average KS3 scores	-0.008**	-0.003	-0.037***	-0.054***
	(0.004)	(0.007)	(0.007)	(0.012)
Observations	7938	7938	6402	6402

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Included but not shown: Specification 3 variables.

Appendix B:

B.1 Descriptive statistics

Table B.1.1: Descriptive Statistics for Control Variables

Specification	Control Variable	Mean	Std. Dev.
Specification 1	Female	0.509	0.500
	White	0.649	0.477
	KS2 score (standardised)	0	1
	Parents' Employment Status		
	MP Employed	0.643	0.479
	MP Self Employed	0.048	0.215
	MP Unemployed	0.024	0.154
	MP Out of the labour	0.272	0.445
	MP Missing	0.012	0.108
	SP Employed	0.479	0.500
	SP Self Employed	0.074	0.263
	SP Unemployed	0.018	0.132
	SP Out of the labour	0.100	0.300
	SP Missing	0.329	0.470
	Parents' Education		
	MP Academic Degree	0.238	0.426
	MP A levels	0.136	0.343
	MP Other Qualification	0.358	0.480
	MP No qualification	0.233	0.423
	MP Missing	0.035	0.183
	SP Academic Degree	0.149	0.356
	SP A levels	0.105	0.307
	SP Other Qualification	0.204	0.403
	SP No qualification	0.191	0.393
	SP Missing	0.351	0.477
Specification 2	*Specification 1 plus ...*		
	MP's Age	42.320	6.361
	SP's Age	43.703	5.705
	Number of siblings	2.027	1.491
	MP Married	0.758	0.428
	MP Single	0.068	0.252
	MP Divorced/ Other	0.171	0.376
	MP Missing	0.003	0.057
	SP Married	0.737	0.440
	SP Single	0.004	0.063
	SP Divorced/ Other	0.005	0.073
	SP Missing	0.253	0.435
Specification 3	*Specification 2 plus ...*		
	Geographic Region		
	London	0.205	0.404
	North East	0.042	0.200
	North West	0.133	0.340
	Yorkshire and Humber	0.105	0.307
	East Midlands	0.074	0.262
	West Midlands	0.262	0.440
	East of England	0.178	0.383
	Class size	22.037	2.051
	Observations	7938	

B.2 First Stage Results

Table B.2.1: First stage estimates and F-statistics

	A1: Stay/Leave	F-statistic	A2: Acad/Voc	F-statistic	A3: Stay/Leave	F -statistic	A4: Acad/Voc	F-statistic
Endogenous variable: Average peers' achievements in KS3 exams								
Instrumental variable: Average peers-of-peers' achievements in KS2 exams								
Specification 1	0.951***	317.00	0.968***	321.35	0.951***	317.00	0.971***	312.63
	(0.054)		(0.054)		(0.054)		(0.055)	
Specification 2	0.951***	309.07	0.968***	312.26	0.951***	309.07	0.971***	305.80
	(0.054)		(0.054)		(0.054)		(0.055)	
Specification 3	0.951***	315.56	0.968***	323.71	0.951***	315.56	0.971***	313.52
	(0.054)		(0.054)		(0.054)		(0.055)	
Endogenous variable: Average peers' aspirations								
Instrumental variable (1): % of peers-of-peers from the highest SEG (2): Average peers-of-peers' achievements in KS2 exams								
Specification 1		37.76		27.61		32.99		28.74
Instrumental variable 1:	0.036***		0.054***		0.026***		0.071***	
	(0.007)		(0.015)		(0.006)		(0.015)	
Instrumental variable 2:	0.063**		0.215***		0.081***		0.145**	
	(0.026)		(0.060)		(0.026)		(0.068)	
Specification 2		36.91		27.61		32.29		28.25
Instrumental variable 1:	0.036***		0.054***		0.026***		0.071***	
	(0.007)		(0.015)		(0.006)		(0.015)	
Instrumental variable 2:	0.063**		0.215***		0.081***		0.145**	
	(0.026)		(0.060)		(0.026)		(0.068)	
Specification 3		33.03		25.41		25.77		24.95
Instrumental variable 1:	0.036***		0.054***		0.026***		0.071***	
	(0.007)		(0.015)		(0.006)		(0.015)	
Instrumental variable 2:	0.063**		0.215***		0.081***		0.145**	
	(0.026)		(0.060)		(0.026)		(0.068)	
Observations	7938		6950		7938		7053	

Robust standard errors clustered at the secondary school level reported in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Angrist-Pischke first-stage F-statistics for strong instruments reported for each first-stage regression (Angrist and Pischke, 2008).
Aspiration A1 (preliminary) and A3 (later): Stay or Leave FT education. Aspiration A2 (preliminary) and A4 (later): Academic or Vocational route.

B.3 Main Results - Complete Tables

B.3.1 Peer effect: Average peers' achievements on KS3 exams

Table B.3.1.1: LPM and IV-2SLS estimates of peer effects on age 14 aspiration to stay/leave FT education

	LPM			IV-2SLS		
	Spec 1	Spec 2	Sepc 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' KS3 score	0.004	0.003	0.003	0.005	0.004	0.004
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.007)
KS2 score	0.073***	0.072***	0.072***	0.073***	0.072***	0.071***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
MP A levels	-0.014	-0.013	-0.013	-0.014	-0.013	-0.013
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
MP Other Qualification	-0.035***	-0.034***	-0.033***	-0.034***	-0.034***	-0.033***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
MP No Qualification	-0.031***	-0.032***	-0.032***	-0.031***	-0.032***	-0.031***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
MP Missing	0.005	0.005	0.006	0.005	0.005	0.005
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
SP A levels	-0.009	-0.008	-0.009	-0.009	-0.008	-0.009
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
SP Other Qualification	-0.014	-0.013	-0.013	-0.013	-0.013	-0.013
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
SP No Qualification	-0.019	-0.017	-0.016	-0.019	-0.016	-0.016
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
SP Missing	-0.004	-0.002	-0.003	-0.004	-0.002	-0.003
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)
MP Self Employed	0.013	0.012	0.012	0.013	0.012	0.012
	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)
MP Unemployed	0.003	0.005	0.005	0.003	0.005	0.005
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
MP Out of the labour	0.003	0.008	0.007	0.004	0.008	0.007
	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
MP Missing	-0.003	-0.010	-0.007	-0.003	-0.009	-0.007
	(0.030)	(0.031)	(0.030)	(0.030)	(0.031)	(0.031)
SP Self Employed	-0.025*	-0.024*	-0.024*	-0.025*	-0.024*	-0.024*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
SP Unemployed	-0.007	-0.005	-0.008	-0.007	-0.005	-0.008
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
SP Out of the labour	-0.004	-0.003	-0.003	-0.004	-0.002	-0.002
	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
SP Missing	-0.026	-0.009	-0.009	-0.025	-0.009	-0.009
	(0.019)	(0.021)	(0.021)	(0.019)	(0.021)	(0.021)
Female	0.067***	0.068***	0.068***	0.067***	0.068***	0.068***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
White	-0.128***	-0.131***	-0.124***	-0.128***	-0.131***	-0.124***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)
MP's Age		0.001***	0.001**		0.001***	0.001**
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		-0.001	-0.001		-0.001	-0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.005**	-0.005**		-0.005**	-0.005**
		(0.002)	(0.002)		(0.002)	(0.002)
MP Single		-0.007	-0.009		-0.007	-0.009
		(0.025)	(0.025)		(0.025)	(0.025)
MP Divorced/ Other		0.024	0.023		0.024	0.023
		(0.024)	(0.024)		(0.024)	(0.024)
MP Missing		-0.003	-0.007		-0.003	-0.007
		(0.058)	(0.058)		(0.058)	(0.058)
SP Single		0.007	0.008		0.007	0.008
		(0.056)	(0.056)		(0.056)	(0.056)
SP Divorced/ Other		0.087***	0.084***		0.087***	0.084***
		(0.031)	(0.031)		(0.031)	(0.031)
SP Missing		-0.038	-0.038		-0.038	-0.038
		(0.024)	(0.024)		(0.024)	(0.024)
North East			-0.030*			-0.030*
			(0.018)			(0.018)
North West			-0.020*			-0.020*
			(0.012)			(0.012)
Yorkshire and Humber			-0.021*			-0.021*
			(0.013)			(0.013)
East Midlands			-0.042***			-0.042***
			(0.016)			(0.016)
West Midlands			-0.020**			-0.020**
			(0.009)			(0.009)
East of England			-0.015			-0.015
			(0.010)			(0.010)
Class size			0.000			0.000
			(0.002)			(0.002)
Constant	0.985***	0.968***	0.987***	0.984***	0.969***	0.986***
	(0.009)	(0.029)	(0.049)	(0.010)	(0.029)	(0.050)
Observations	7938	7938	7938	7938	7938	7938

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Table B.3.1.2: LPM and IV-2SLS estimates of peer effects on age 14 aspiration to follow academic/vocational route

	LPM			IV-2SLS		
	Spec 1	Spec 2	Spec 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' KS3 score	0.055***	0.052***	0.044***	0.050***	0.045***	0.039***
	(0.008)	(0.008)	(0.008)	(0.014)	(0.014)	(0.014)
KS2 score	0.065***	0.062***	0.062***	0.067***	0.065***	0.064***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
MP A levels	-0.057***	-0.055***	-0.051***	-0.059***	-0.056***	-0.052***
	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
MP Other Qualification	-0.065***	-0.063***	-0.061***	-0.067***	-0.065***	-0.063***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
MP No Qualification	-0.050***	-0.051***	-0.047**	-0.053***	-0.054***	-0.050***
	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)
MP Missing	-0.006	-0.003	-0.003	-0.006	-0.002	-0.002
	(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.030)
SP A levels	0.010	0.013	0.014	0.008	0.012	0.013
	(0.019)	(0.019)	(0.018)	(0.019)	(0.019)	(0.018)
SP Other Qualification	-0.017	-0.012	-0.010	-0.020	-0.015	-0.013
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
SP No Qualification	-0.031	-0.026	-0.025	-0.034*	-0.029	-0.027
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
SP Missing	-0.006	-0.002	0.000	-0.006	-0.003	-0.000
	(0.041)	(0.040)	(0.038)	(0.040)	(0.040)	(0.038)
MP Self Employed	-0.011	-0.013	-0.016	-0.011	-0.012	-0.015
	(0.025)	(0.025)	(0.024)	(0.025)	(0.025)	(0.024)
MP Unemployed	-0.061	-0.050	-0.036	-0.062	-0.051	-0.036
	(0.039)	(0.040)	(0.038)	(0.039)	(0.040)	(0.037)
MP Out of the labour	-0.004	0.006	0.010	-0.005	0.005	0.009
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
MP Missing	-0.028	-0.042	-0.048	-0.034	-0.049	-0.053
	(0.060)	(0.060)	(0.059)	(0.060)	(0.060)	(0.060)
SP Self Employed	0.010	0.009	0.004	0.011	0.010	0.005
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
SP Unemployed	0.064	0.069	0.083*	0.061	0.066	0.081*
	(0.044)	(0.044)	(0.045)	(0.044)	(0.044)	(0.045)
SP Out of the labour	-0.006	0.001	-0.001	-0.009	-0.001	-0.003
	(0.023)	(0.023)	(0.022)	(0.023)	(0.022)	(0.022)
SP Missing	-0.051	-0.024	-0.027	-0.054	-0.027	-0.029
	(0.041)	(0.043)	(0.042)	(0.041)	(0.043)	(0.042)
Female	-0.000	0.001	0.002	-0.000	0.001	0.002
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
White	-0.091***	-0.097***	-0.103***	-0.090***	-0.095***	-0.102***
	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)	(0.016)
MP's Age		0.001	0.002*		0.001	0.002*
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.013***	-0.013***		-0.013***	-0.014***
		(0.004)	(0.004)		(0.004)	(0.004)
MP Single		-0.112***	-0.095**		-0.113***	-0.095**
		(0.040)	(0.040)		(0.040)	(0.040)
MP Divorced/ Other		-0.054	-0.047		-0.053	-0.046
		(0.036)	(0.036)		(0.036)	(0.036)
MP Missing		0.006	0.040		0.006	0.039
		(0.093)	(0.092)		(0.092)	(0.091)
SP Single		-0.091	-0.077		-0.092	-0.078
		(0.096)	(0.097)		(0.095)	(0.097)
SP Divorced/ Other		-0.034	-0.050		-0.036	-0.051
		(0.078)	(0.077)		(0.078)	(0.077)
SP Missing		0.026	0.022		0.025	0.020
		(0.039)	(0.039)		(0.039)	(0.039)
North East			0.041			0.039
			(0.043)			(0.043)
North West			-0.034			-0.034
			(0.036)			(0.036)
Yorkshire and Humber			0.041			0.038
			(0.033)			(0.033)
East Midlands			0.041			0.043
			(0.036)			(0.036)
West Midlands			-0.003			-0.004
			(0.024)			(0.024)
East of England			0.037			0.036
			(0.025)			(0.025)
Class size			-0.031***			-0.032***
			(0.004)			(0.004)
Constant	0.851***	0.768***	1.431***	0.856***	0.767***	1.444***
	(0.018)	(0.050)	(0.108)	(0.019)	(0.050)	(0.109)
Observations	6950	6950	6950	6950	6950	6950

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Table B.3.1.3: LPM and IV-2SLS estimates of peer effects on age 16 aspiration to stay/leave FT education

	LPM			IV-2SLS		
	Spec 1	Spec 2	Spec 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' KS3 scores	0.012***	0.009***	0.009**	0.011*	0.007	0.006
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)
KS2 score	0.057***	0.054***	0.054***	0.057***	0.055***	0.054***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
MP A levels	-0.031***	-0.028***	-0.029***	-0.031***	-0.029***	-0.029***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
MP Other Qualification	-0.045***	-0.043***	-0.043***	-0.046***	-0.044***	-0.043***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
MP No Qualification	-0.053***	-0.053***	-0.053***	-0.053***	-0.053***	-0.053***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
MP Missing	-0.017	-0.018	-0.018	-0.017	-0.018	-0.018
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
SP A levels	-0.011	-0.008	-0.009	-0.011	-0.009	-0.009
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
SP Other Qualification	-0.030***	-0.025**	-0.025**	-0.030***	-0.026***	-0.025**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
SP No Qualification	-0.042***	-0.039***	-0.039***	-0.042***	-0.040***	-0.040***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
SP Missing	0.004	0.006	0.005	0.004	0.006	0.005
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
MP Self Employed	0.002	0.002	0.003	0.002	0.002	0.003
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
MP Unemployed	-0.012	-0.006	-0.007	-0.012	-0.006	-0.007
	(0.023)	(0.023)	(0.022)	(0.023)	(0.023)	(0.022)
MP Out of the labour	0.021***	0.025***	0.024***	0.021***	0.025***	0.024***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
MP Missing	0.018	0.021	0.027	0.018	0.020	0.026
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
SP Self Employed	-0.009	-0.010	-0.010	-0.009	-0.010	-0.010
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
SP Unemployed	0.012	0.014	0.010	0.012	0.014	0.010
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
SP Out of the labour	0.026**	0.028**	0.031**	0.026**	0.028**	0.030**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
SP Missing	-0.032	-0.026	-0.026	-0.032	-0.026	-0.026
	(0.020)	(0.021)	(0.021)	(0.019)	(0.020)	(0.021)
Female	0.076***	0.076***	0.076***	0.076***	0.076***	0.076***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
White	-0.125***	-0.129***	-0.119***	-0.125***	-0.128***	-0.118***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)
MP's Age		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		0.002***	0.002**		0.002***	0.002***
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.007***	-0.007***		-0.007***	-0.007***
		(0.002)	(0.002)		(0.002)	(0.002)
MP Single		0.002	-0.000		0.002	-0.000
		(0.026)	(0.026)		(0.026)	(0.026)
MP Divorced/ Other		0.041*	0.040*		0.041*	0.041*
		(0.025)	(0.024)		(0.025)	(0.024)
MP Missing		0.001	-0.008		0.001	-0.008
		(0.055)	(0.056)		(0.055)	(0.055)
SP Single		-0.107	-0.105		-0.107	-0.104
		(0.072)	(0.071)		(0.072)	(0.071)
SP Divorced/ Other		0.033	0.031		0.033	0.030
		(0.046)	(0.045)		(0.045)	(0.045)
SP Missing		-0.034	-0.033		-0.034	-0.033
		(0.025)	(0.025)		(0.025)	(0.025)
North East			-0.049**			-0.050**
			(0.022)			(0.022)
North West			-0.030**			-0.030**
			(0.012)			(0.012)
Yorkshire and Humber			-0.044***			-0.044***
			(0.014)			(0.014)
East Midlands			-0.027*			-0.027
			(0.016)			(0.016)
West Midlands			-0.012			-0.012
			(0.008)			(0.008)
East of England			-0.029**			-0.029**
			(0.012)			(0.011)
Class size			0.001			0.001
			(0.002)			(0.002)
Constant	0.989***	0.891***	0.897***	0.989***	0.890***	0.899***
	(0.010)	(0.032)	(0.050)	(0.010)	(0.032)	(0.051)
Observations	7938	7938	7938	7938	7938	7938

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Table B.3.1.4: LPM and IV-2SLS estimates of peer effects on age 16 aspiration to follow academic/vocational route

	LPM			IV-2SLS		
	Spec 1	Spec 2	Spec 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers KS3 scores	0.036***	0.033***	0.025***	0.040***	0.036**	0.029**
	(0.008)	(0.008)	(0.007)	(0.014)	(0.014)	(0.013)
KS2 score	0.134***	0.130***	0.128***	0.134***	0.130***	0.128***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
MP A levels	-0.045***	-0.041***	-0.036**	-0.045***	-0.041***	-0.036**
	(0.016)	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)
MP Other Qualification	-0.069***	-0.065***	-0.061***	-0.069***	-0.065***	-0.061***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
MP No Qualification	-0.040**	-0.047**	-0.042**	-0.040**	-0.046**	-0.042**
	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.018)
MP Missing	-0.037	-0.037	-0.034	-0.037	-0.037	-0.035
	(0.040)	(0.040)	(0.038)	(0.040)	(0.040)	(0.038)
SP A levels	-0.021	-0.019	-0.019	-0.021	-0.018	-0.019
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.017)
SP Other Qualification	-0.029*	-0.023	-0.024	-0.029*	-0.023	-0.023
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
SP No Qualification	-0.062***	-0.058***	-0.059***	-0.062***	-0.057***	-0.059***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
SP Missing	-0.041	-0.036	-0.033	-0.041	-0.036	-0.033
	(0.039)	(0.039)	(0.038)	(0.039)	(0.039)	(0.038)
MP Self Employed	0.020	0.014	0.014	0.020	0.013	0.014
	(0.024)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
MP Unemployed	-0.007	0.001	0.013	-0.007	0.001	0.013
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
MP Out of the labour	-0.011	-0.002	-0.001	-0.011	-0.002	-0.001
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
MP Missing	0.137**	0.106*	0.113**	0.138**	0.107*	0.114**
	(0.057)	(0.057)	(0.056)	(0.057)	(0.056)	(0.055)
SP Self Employed	-0.002	-0.002	-0.005	-0.002	-0.002	-0.005
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)
SP Unemployed	-0.057	-0.054	-0.052	-0.057	-0.053	-0.052
	(0.044)	(0.044)	(0.043)	(0.044)	(0.044)	(0.043)
SP Out of the labour	0.045**	0.042**	0.044**	0.046**	0.042**	0.044**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
SP Missing	-0.066	0.008	0.003	-0.066	0.009	0.003
	(0.041)	(0.044)	(0.043)	(0.040)	(0.043)	(0.043)
Female	0.028**	0.030**	0.030***	0.028**	0.030**	0.030***
	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)
White	-0.184***	-0.187***	-0.164***	-0.184***	-0.187***	-0.164***
	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)	(0.016)
MP's Age		0.003***	0.004***		0.003***	0.004***
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		0.000	-0.000		0.000	-0.000
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.006	-0.005		-0.006	-0.005
		(0.004)	(0.004)		(0.004)	(0.004)
MP Single		-0.100**	-0.090**		-0.100**	-0.090**
		(0.042)	(0.042)		(0.042)	(0.042)
MP Divorced/ Other		-0.014	-0.008		-0.014	-0.008
		(0.039)	(0.039)		(0.038)	(0.038)
MP Missing		-0.021	-0.007		-0.020	-0.007
		(0.088)	(0.086)		(0.088)	(0.086)
SP Single		0.010	0.018		0.010	0.018
		(0.098)	(0.096)		(0.098)	(0.096)
SP Divorced/ Other		-0.014	-0.041		-0.014	-0.041
		(0.080)	(0.083)		(0.080)	(0.082)
SP Missing		-0.069*	-0.073*		-0.069*	-0.073*
		(0.038)	(0.038)		(0.038)	(0.038)
North East			-0.048			-0.047
			(0.042)			(0.042)
North West			-0.107***			-0.107***
			(0.033)			(0.033)
Yorkshire and Humber			-0.052*			-0.052*
			(0.029)			(0.028)
East Midlands			-0.059*			-0.059*
			(0.032)			(0.032)
West Midlands			-0.057**			-0.057**
			(0.024)			(0.024)
East of England			-0.057**			-0.057**
			(0.028)			(0.028)
Class size			-0.029***			-0.029***
			(0.004)			(0.004)
Constant	0.864***	0.727***	1.404***	0.863***	0.727***	1.401***
	(0.020)	(0.051)	(0.103)	(0.021)	(0.051)	(0.105)
Observations	7053	7053	7053	7053	7053	7053

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

B.3.2 Peer effect: Average peers' educational aspirations

Table B.3.2.1: LPM and IV-2SLS estimates of peer effects on age 14 aspiration to stay/leave FT education

	LPM			IV-2SLS		
	Spec 1	Spec 2	Spec 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Aspiration 1	0.220***	0.220***	0.212***	0.159	0.128	0.131
	(0.043)	(0.043)	(0.045)	(0.120)	(0.124)	(0.130)
KS2 score	0.071***	0.069***	0.070***	0.072***	0.071***	0.071***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
MP A levels	-0.013	-0.012	-0.012	-0.013	-0.013	-0.012
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
MP Other Qualification	-0.032***	-0.031***	-0.031***	-0.033***	-0.033***	-0.032***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
MP No Qualification	-0.031***	-0.032***	-0.032***	-0.032***	-0.033***	-0.032***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
MP Missing	0.008	0.007	0.008	0.007	0.006	0.007
	(0.020)	(0.021)	(0.020)	(0.020)	(0.021)	(0.021)
SP A levels	-0.008	-0.008	-0.008	-0.009	-0.008	-0.009
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
SP Other Qualification	-0.011	-0.011	-0.011	-0.012	-0.012	-0.012
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
SP No Qualification	-0.018	-0.015	-0.015	-0.019	-0.016	-0.016
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
SP Missing	-0.002	-0.001	-0.001	-0.003	-0.001	-0.001
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
MP Self Employed	0.013	0.011	0.011	0.013	0.012	0.012
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
MP Unemployed	0.001	0.003	0.003	0.001	0.003	0.003
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
MP Out of the labour	0.003	0.008	0.008	0.003	0.008	0.007
	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
MP Missing	-0.007	-0.013	-0.012	-0.007	-0.012	-0.011
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)
SP Self Employed	-0.025*	-0.023*	-0.024*	-0.025*	-0.023*	-0.024*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
SP Unemployed	-0.011	-0.008	-0.010	-0.010	-0.007	-0.009
	(0.024)	(0.024)	(0.024)	(0.025)	(0.025)	(0.024)
SP Out of the labour	-0.007	-0.005	-0.005	-0.006	-0.004	-0.004
	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
SP Missing	-0.026	-0.010	-0.010	-0.026	-0.010	-0.010
	(0.019)	(0.021)	(0.021)	(0.019)	(0.021)	(0.021)
Female	0.066***	0.066***	0.066***	0.066***	0.067***	0.067***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
White	-0.113***	-0.116***	-0.115***	-0.117***	-0.122***	-0.118***
	(0.008)	(0.008)	(0.008)	(0.010)	(0.011)	(0.010)
MP's Age		0.001**	0.001**		0.001**	0.001**
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		-0.001	-0.001		-0.001	-0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.005**	-0.006**		-0.005**	-0.005**
		(0.002)	(0.002)		(0.002)	(0.002)
MP Single		-0.008	-0.010		-0.008	-0.010
		(0.025)	(0.025)		(0.025)	(0.025)
MP Divorced/ Other		0.022	0.021		0.023	0.022
		(0.024)	(0.024)		(0.024)	(0.024)
MP Missing		-0.007	-0.007		-0.005	-0.007
		(0.058)	(0.058)		(0.058)	(0.058)
SP Single		0.015	0.015		0.012	0.013
		(0.055)	(0.054)		(0.055)	(0.055)
SP Divorced/ Other		0.090***	0.089***		0.088***	0.087***
		(0.031)	(0.031)		(0.031)	(0.032)
SP Missing		-0.036	-0.036		-0.037	-0.036
		(0.024)	(0.024)		(0.024)	(0.024)
North East			-0.014			-0.020
			(0.016)			(0.020)
North West			-0.003			-0.009
			(0.011)			(0.015)
Yorkshire and Humber			-0.007			-0.013
			(0.011)			(0.014)
East Midlands			-0.026*			-0.031*
			(0.014)			(0.017)
West Midlands			-0.006			-0.011
			(0.008)			(0.012)
East of England			-0.001			-0.006
			(0.010)			(0.013)
Class size			0.000			0.000
			(0.001)			(0.002)
Constant	0.779***	0.775***	0.780***	0.837***	0.854***	0.858***
	(0.041)	(0.048)	(0.065)	(0.111)	(0.111)	(0.137)
Observations	7938	7938	7938	7938	7938	7938

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Table B.3.2.2: LPM and IV-2SLS estimates of peer effects on age 14 aspiration to follow academic/vocational route

	LPM			IV-2SLS		
	Spec 1	Spec 2	Spec 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Aspiration 2	0.747***	0.741***	0.707***	0.641***	0.592***	0.551***
	(0.021)	(0.021)	(0.023)	(0.081)	(0.085)	(0.095)
KS2 score	0.060***	0.056***	0.057***	0.063***	0.060***	0.061***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
MP A levels	-0.051***	-0.048***	-0.048***	-0.053***	-0.050***	-0.049***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
MP Other Qualification	-0.052***	-0.050***	-0.050***	-0.056***	-0.055***	-0.055***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
MP No Qualification	-0.041**	-0.040**	-0.040**	-0.045***	-0.046***	-0.045**
	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)
MP Missing	-0.016	-0.014	-0.013	-0.014	-0.010	-0.010
	(0.029)	(0.029)	(0.029)	(0.029)	(0.028)	(0.029)
SP A levels	0.012	0.015	0.015	0.011	0.014	0.014
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
SP Other Qualification	-0.018	-0.013	-0.013	-0.021	-0.016	-0.015
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
SP No Qualification	-0.030	-0.025	-0.025	-0.032*	-0.028	-0.027
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
SP Missing	-0.003	-0.000	0.001	-0.004	-0.001	0.000
	(0.032)	(0.032)	(0.032)	(0.033)	(0.033)	(0.033)
MP Self Employed	-0.006	-0.007	-0.009	-0.006	-0.007	-0.010
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
MP Unemployed	-0.046	-0.036	-0.032	-0.049	-0.040	-0.033
	(0.033)	(0.034)	(0.033)	(0.034)	(0.035)	(0.034)
MP Out of the labour	0.005	0.015	0.016	0.003	0.012	0.014
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
MP Missing	-0.019	-0.024	-0.029	-0.026	-0.034	-0.039
	(0.053)	(0.053)	(0.053)	(0.053)	(0.054)	(0.054)
SP Self Employed	0.008	0.007	0.006	0.009	0.008	0.006
	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)	(0.017)
SP Unemployed	0.054	0.060	0.067	0.054	0.059	0.069
	(0.046)	(0.046)	(0.046)	(0.045)	(0.045)	(0.045)
SP Out of the labour	-0.009	-0.001	-0.003	-0.011	-0.003	-0.004
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
SP Missing	-0.052	-0.043	-0.044	-0.055*	-0.042	-0.042
	(0.032)	(0.034)	(0.034)	(0.033)	(0.035)	(0.035)
Female	0.004	0.005	0.005	0.003	0.004	0.005
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
White	-0.054***	-0.061***	-0.072***	-0.059***	-0.068***	-0.078***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)
MP's Age		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.013***	-0.013***		-0.013***	-0.013***
		(0.004)	(0.004)		(0.004)	(0.004)
MP Single		-0.087**	-0.081**		-0.094**	-0.086**
		(0.038)	(0.039)		(0.039)	(0.039)
MP Divorced/ Other		-0.030	-0.029		-0.034	-0.032
		(0.034)	(0.034)		(0.034)	(0.035)
MP Missing		-0.004	0.010		-0.002	0.017
		(0.088)	(0.089)		(0.087)	(0.087)
SP Single		-0.079	-0.076		-0.083	-0.077
		(0.081)	(0.083)		(0.084)	(0.085)
SP Divorced/ Other		-0.052	-0.053		-0.050	-0.054
		(0.075)	(0.075)		(0.075)	(0.075)
SP Missing		0.028	0.028		0.026	0.025
		(0.036)	(0.037)		(0.037)	(0.037)
North East			0.046**			0.043*
			(0.020)			(0.024)
North West			0.012			0.000
			(0.014)			(0.020)
Yorkshire and Humber			0.028*			0.028
			(0.016)			(0.019)
East Midlands			0.027*			0.032
			(0.015)			(0.020)
West Midlands			0.023*			0.016
			(0.012)			(0.015)
East of England			0.031**			0.031**
			(0.013)			(0.015)
Class size			-0.009***			-0.015***
			(0.002)			(0.004)
Constant	0.275***	0.207***	0.413***	0.373***	0.331***	0.681***
	(0.022)	(0.049)	(0.070)	(0.068)	(0.077)	(0.163)
Observations	6950	6950	6950	6950	6950	6950

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Table B.3.2.3: LPM and IV-2SLS estimates of peer effects on age 16 aspiration to stay/leave FT education

	LPM			IV-2SLS		
	Spec 1	Spec 2	Spec 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Aspiration 3	0.249***	0.237***	0.213***	0.345***	0.264**	0.232*
	(0.044)	(0.044)	(0.048)	(0.117)	(0.123)	(0.137)
KS2 score	0.057***	0.054***	0.054***	0.055***	0.054***	0.054***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
MP A levels	-0.030***	-0.028***	-0.028***	-0.029***	-0.028***	-0.028***
	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)
MP Other Qualification	-0.043***	-0.041***	-0.041***	-0.041***	-0.040***	-0.040***
	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
MP No Qualification	-0.053***	-0.053***	-0.053***	-0.052***	-0.052***	-0.052***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
MP Missing	-0.015	-0.016	-0.017	-0.015	-0.016	-0.016
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
SP A levels	-0.012	-0.009	-0.009	-0.011	-0.008	-0.009
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
SP Other Qualification	-0.030***	-0.025**	-0.024**	-0.028***	-0.024**	-0.024**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
SP No Qualification	-0.042***	-0.039***	-0.039***	-0.041***	-0.039***	-0.039***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
SP Missing	0.006	0.007	0.006	0.006	0.007	0.006
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
MP Self Employed	0.002	0.002	0.002	0.002	0.001	0.002
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
MP Unemployed	-0.014	-0.009	-0.009	-0.015	-0.009	-0.010
	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
MP Out of the labour	0.020***	0.024***	0.024***	0.020***	0.025***	0.024***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
MP Missing	0.011	0.015	0.020	0.010	0.015	0.019
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
SP Self Employed	-0.007	-0.008	-0.008	-0.006	-0.008	-0.008
	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)
SP Unemployed	0.006	0.008	0.007	0.005	0.008	0.007
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
SP Out of the labour	0.024**	0.027**	0.029**	0.025**	0.027**	0.029**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
SP Missing	-0.034*	-0.029	-0.029	-0.034*	-0.030	-0.029
	(0.019)	(0.020)	(0.021)	(0.019)	(0.020)	(0.020)
Female	0.075***	0.075***	0.075***	0.074***	0.075***	0.075***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
White	-0.107***	-0.111***	-0.108***	-0.099***	-0.109***	-0.107***
	(0.008)	(0.008)	(0.008)	(0.011)	(0.012)	(0.010)
MP's Age		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		0.002***	0.002***		0.002***	0.002**
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.007***	-0.007***		-0.007***	-0.007***
		(0.002)	(0.002)		(0.002)	(0.002)
MP Single		0.000	-0.001		0.000	-0.001
		(0.026)	(0.026)		(0.026)	(0.026)
MP Divorced/ Other		0.041*	0.040*		0.041*	0.040*
		(0.025)	(0.024)		(0.025)	(0.024)
MP Missing		-0.002	-0.007		-0.003	-0.007
		(0.054)	(0.054)		(0.054)	(0.054)
SP Single		-0.103	-0.102		-0.102	-0.101
		(0.071)	(0.071)		(0.070)	(0.070)
SP Divorced/ Other		0.034	0.033		0.034	0.033
		(0.045)	(0.045)		(0.045)	(0.044)
SP Missing		-0.032	-0.031		-0.032	-0.031
		(0.025)	(0.025)		(0.025)	(0.025)
North East			-0.026			-0.022
			(0.019)			(0.024)
North West			-0.014			-0.012
			(0.010)			(0.014)
Yorkshire and Humber			-0.028**			-0.025
			(0.012)			(0.016)
East Midlands			-0.010			-0.008
			(0.014)			(0.017)
West Midlands			-0.002			-0.000
			(0.007)			(0.010)
East of England			-0.015			-0.012
			(0.010)			(0.013)
Class size			0.001			0.001
			(0.001)			(0.001)
Constant	0.755***	0.674***	0.691***	0.652***	0.638***	0.659***
	(0.041)	(0.050)	(0.067)	(0.113)	(0.115)	(0.147)
Observations	7938	7938	7938	7938	7938	7938

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Table B.3.2.4: LPM and IV-2SLS estimates of peer effects on age 16 aspiration to follow academic/vocational route

	LPM			IV-2SLS		
	Spec 1	Spec 2	Spec 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Aspiration 4	0.633***	0.626***	0.592***	0.441***	0.400***	0.340***
	(0.027)	(0.027)	(0.029)	(0.090)	(0.094)	(0.108)
KS2 score	0.124***	0.120***	0.120***	0.130***	0.127***	0.126***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
MP A levels	-0.036**	-0.032**	-0.031**	-0.040***	-0.037**	-0.035**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
MP Other Qualification	-0.055***	-0.051***	-0.050***	-0.062***	-0.059***	-0.057***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
MP No Qualification	-0.035**	-0.039**	-0.038**	-0.041**	-0.046***	-0.043**
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
MP Missing	-0.011	-0.011	-0.012	-0.018	-0.020	-0.021
	(0.032)	(0.033)	(0.032)	(0.034)	(0.034)	(0.034)
SP A levels	-0.025	-0.023	-0.023	-0.025	-0.022	-0.022
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
SP Other Qualification	-0.024	-0.020	-0.020	-0.029*	-0.024	-0.024
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
SP No Qualification	-0.056***	-0.052***	-0.052***	-0.061***	-0.057***	-0.058***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
SP Missing	-0.023	-0.018	-0.018	-0.028	-0.025	-0.024
	(0.033)	(0.033)	(0.033)	(0.034)	(0.034)	(0.035)
MP Self Employed	0.028	0.022	0.021	0.026	0.019	0.018
	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
MP Unemployed	-0.004	0.004	0.008	-0.006	0.002	0.010
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.035)
MP Out of the labour	-0.006	0.002	0.003	-0.009	-0.000	-0.000
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
MP Missing	0.120**	0.094*	0.096*	0.119**	0.091*	0.098*
	(0.051)	(0.051)	(0.051)	(0.052)	(0.052)	(0.053)
SP Self Employed	-0.002	-0.002	-0.003	-0.001	-0.001	-0.003
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
SP Unemployed	-0.068*	-0.064	-0.061	-0.068	-0.063	-0.059
	(0.041)	(0.041)	(0.041)	(0.042)	(0.042)	(0.042)
SP Out of the labour	0.042**	0.040**	0.040**	0.040**	0.038**	0.040**
	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
SP Missing	-0.077**	-0.011	-0.012	-0.077**	-0.007	-0.008
	(0.033)	(0.036)	(0.036)	(0.034)	(0.038)	(0.038)
Female	0.028***	0.029***	0.030***	0.028***	0.030***	0.030***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)
White	-0.122***	-0.126***	-0.126***	-0.140***	-0.147***	-0.142***
	(0.012)	(0.012)	(0.013)	(0.015)	(0.015)	(0.016)
MP's Age		0.003***	0.003***		0.003***	0.003***
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		-0.000	-0.000		-0.000	-0.000
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.006	-0.006		-0.007*	-0.006
		(0.004)	(0.004)		(0.004)	(0.004)
MP Single		-0.088**	-0.085**		-0.094**	-0.088**
		(0.041)	(0.041)		(0.041)	(0.041)
MP Divorced/ Other		-0.013	-0.012		-0.012	-0.009
		(0.037)	(0.037)		(0.037)	(0.037)
MP Missing		-0.043	-0.035		-0.036	-0.023
		(0.083)	(0.083)		(0.084)	(0.083)
SP Single		-0.004	0.000		0.000	0.008
		(0.099)	(0.098)		(0.097)	(0.096)
SP Divorced/ Other		-0.059	-0.063		-0.043	-0.055
		(0.077)	(0.078)		(0.077)	(0.079)
SP Missing		-0.061*	-0.062*		-0.065*	-0.068*
		(0.037)	(0.037)		(0.037)	(0.037)
North East			-0.004			-0.026
			(0.020)			(0.032)
North West			-0.022			-0.061**
			(0.017)			(0.027)
Yorkshire and Humber			-0.005			-0.029
			(0.015)			(0.022)
East Midlands			-0.014			-0.032
			(0.016)			(0.023)
West Midlands			-0.006			-0.029
			(0.013)			(0.019)
East of England			-0.006			-0.030
			(0.015)			(0.022)
Class size			-0.010***			-0.019***
			(0.002)			(0.005)
Constant	0.381***	0.280***	0.519***	0.543***	0.448***	0.934***
	(0.029)	(0.052)	(0.073)	(0.073)	(0.084)	(0.193)
Observations	7053	7053	7053	7053	7053	7053

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

B.4 Sensitivity Analyses

Table B.4.1: Sensitivity analysis to excluding observations from small secondary schools

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT Ed		Acad/Voc route		Stay/Leave FT Ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.006	0.003	0.041***	0.033**	0.009**	0.009	0.023***	0.026**
	(0.003)	(0.006)	(0.008)	(0.013)	(0.004)	(0.006)	(0.007)	(0.013)
Peers' Average Aspirations	0.200***	0.158	0.685***	0.514***	0.235***	0.269**	0.584***	0.305**
	(0.046)	(0.127)	(0.025)	(0.111)	(0.049)	(0.136)	(0.031)	(0.119)
Observations	7326	7326	6413	6413	7326	7326	6502	6502

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included but not shown: Specification 3 variables.

Table B.4.2: Sensitivity analysis to excluding schools with less than 10 LSYPE individuals

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT Ed		Acad/Voc route		Stay/Leave FT Ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.003	0.003	0.039***	0.041***	0.010***	0.011*	0.020**	0.023
	(0.004)	(0.007)	(0.008)	(0.014)	(0.004)	(0.007)	(0.008)	(0.014)
Peers' Average Aspirations	0.186***	0.099	0.705***	0.566***	0.191***	0.297**	0.595***	0.243*
	(0.050)	(0.142)	(0.024)	(0.097)	(0.053)	(0.129)	(0.031)	(0.126)
Observations	7366	7366	6446	6446	7366	7366	6540	6540

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included but not shown: Specification 3 variables.

Table B.4.3: Sensitivity analysis to including only individuals who have more than 1 peer of peer

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT Ed		Acad/Voc route		Stay/Leave FT Ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.003	0.002	0.047***	0.040**	0.008**	0.007	0.024***	0.025*
	(0.004)	(0.007)	(0.008)	(0.015)	(0.004)	(0.007)	(0.008)	(0.015)
Peers' Average Aspirations	0.236***	0.067	0.699***	0.546***	0.236***	0.258	0.594***	0.323**
	(0.047)	(0.168)	(0.024)	(0.106)	(0.050)	(0.160)	(0.032)	(0.137)
Observations	7045	7045	6194	6194	7045	7045	6266	6266

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included but not shown: Specification 3 variables.

Table B.4.4: Sensitivity analysis to including only individuals who have 5 or more peers of peers

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT Ed		Acad/Voc route		Stay/Leave FT Ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.004	0.003	0.053***	0.055***	0.008**	0.005	0.026***	0.024
	(0.005)	(0.008)	(0.008)	(0.017)	(0.004)	(0.007)	(0.008)	(0.017)
Peers' Average Aspirations	0.201***	0.062	0.700***	0.656***	0.158***	0.173	0.585***	0.316*
	(0.055)	(0.188)	(0.027)	(0.100)	(0.047)	(0.220)	(0.037)	(0.170)
Observations	5677	5677	5032	5032	5677	5677	5086	5086

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included but not shown: Specification 3 variables.

Table B.4.5: Sensitivity analysis to including only individuals who have 10 or more peers of peers

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT Ed		Acad/Voc route		Stay/Leave FT Ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.008*	0.002	0.051***	0.069***	0.013***	0.007	0.036***	0.040**
	(0.004)	(0.009)	(0.009)	(0.018)	(0.004)	(0.009)	(0.009)	(0.017)
Peers' Average Aspirations	0.182***	0.125	0.674***	0.790***	0.222***	0.417	0.538***	0.594***
	(0.063)	(0.263)	(0.034)	(0.115)	(0.059)	(0.383)	(0.051)	(0.162)
Observations	4023	4023	3642	3642	4023	4023	3663	3663

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included but not shown: Specification 3 variables.

Table B.4.6: Sensitivity analysis to including only individuals who have 15 or more peers of peers

	Preliminary aspirations: age 14				Later aspirations: age 16			
	Stay/Leave FT Ed		Acad/Voc route		Stay/Leave FT Ed		Acad/Voc route	
	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS	LPM	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peers' Average KS3 scores	0.012**	0.010	0.045***	0.062**	0.008	-0.001	0.043***	0.046*
	(0.005)	(0.011)	(0.012)	(0.024)	(0.006)	(0.012)	(0.012)	(0.025)
Peers' Average Aspirations	0.196**	0.390	0.651***	0.606***	0.171**	-0.295	0.559***	0.681***
	(0.077)	(0.348)	(0.042)	(0.135)	(0.071)	(0.856)	(0.063)	(0.241)
Observations	2641	2641	2397	2397	2641	2641	2423	2423

Robust standard errors clustered at the secondary school level are reported in in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included but not shown: Specification 3 variables.

B.5 Analysis on Aspiration 5

The analysis of this chapter was based on the assumption that individuals who had plans to attend a sixth form were aspiring to an academic route while individuals who planned to attend a college of FE or another type of college were aspiring to a vocational route. Although it is true that young individuals who attend a college of FE, rather than a sixth form, tend to be more likely to follow a vocational route as they are exposed to a variety of vocational subjects it is still possible for them to follow an academic route or to continue with academic education in a university after achieving vocational qualifications.

The purpose of this section is to investigate further the effect of peers on the educational aspirations of young individuals at age 15-16 who aspired to stay in FT education after completing the compulsory level, based on what qualifications they aspired to do. The derivation of the aspirations variable used for this analysis is exactly the same as the one used in Chapter 2 with the derivation procedure explained in detail in Appendix A.1. The individuals who mentioned that they aspired to do A levels and/ or attend a university were considered as aspiring to follow an academic route while the individuals who aspired to other qualifications were considered as aspiring to a vocational route. The inclusion of university aspirations in the measure of Aspiration 5 results in individuals who aspired to study for vocational upper-secondary qualifications with the aim to go to university ending up flagged as aspiring the academic route.

The main findings from this analysis are presented in Tables B.5.1-B.5.2 below. As expected, the LPM estimations show positive and significant peer effects on the probability of aspiring to study for academic rather than vocational qualifications for all three specifications and for both peer measures although they are smaller in size than the equivalent coefficients for Aspiration 4. Interestingly, when accounting for potential selection in peer groups instrumenting secondary school peers with the peers-of-peers, the statistical significance of the effect disappears. Perhaps, the influence of peers is stronger in determining where the individuals aspire to go (choosing between a sixth form or a college of FE) rather than what qualifications they aspired to study for. The findings of this analysis show that other variables such as own ability, as estimated from average achievements in KS2 exams, and parents' education are more important predictors of educational aspirations when considering what qualifications to study for.

Table B.5.1: LPM and IV-2SLS estimates of peers' ability on age 16 aspiration to study for academic/vocational qualifications (Aspiration 5)

	LPM			IV-2SLS		
	Spec 1	Spec 2	Sepc 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' KS3 scores	0.017***	0.013**	0.011**	0.007	0.002	-0.000
	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)
KS2 score	0.150***	0.145***	0.144***	0.152***	0.148***	0.146***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
MP A levels	-0.036***	-0.032**	-0.029**	-0.037***	-0.033**	-0.030**
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
MP Other Qualification	-0.073***	-0.069***	-0.067***	-0.076***	-0.071***	-0.069***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
MP No Qualification	-0.069***	-0.072***	-0.072***	-0.073***	-0.076***	-0.075***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
MP Missing	0.019	0.020	0.022	0.020	0.021	0.023
	(0.029)	(0.029)	(0.028)	(0.029)	(0.029)	(0.028)
SP A levels	-0.026*	-0.022	-0.022	-0.028*	-0.023*	-0.023*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
SP Other Qualification	-0.044***	-0.037***	-0.036***	-0.046***	-0.039***	-0.039***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
SP No Qualification	-0.037**	-0.031**	-0.032**	-0.040**	-0.034**	-0.035**
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
SP Missing	-0.016	-0.011	-0.009	-0.017	-0.011	-0.010
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
MP Self Employed	0.024	0.020	0.022	0.024	0.020	0.022
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
MP Unemployed	0.000	0.011	0.014	-0.001	0.009	0.013
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
MP Out of the labour	0.008	0.018*	0.018*	0.007	0.017*	0.016
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
MP Missing	0.047	0.035	0.041	0.041	0.029	0.035
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
SP Self Employed	-0.009	-0.010	-0.011	-0.008	-0.009	-0.010
	(0.016)	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)
SP Unemployed	-0.012	-0.009	-0.014	-0.015	-0.012	-0.016
	(0.034)	(0.034)	(0.034)	(0.034)	(0.033)	(0.034)
SP Out of the labour	0.019	0.021	0.021	0.017	0.018	0.019
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
SP Missing	-0.061*	-0.020	-0.022	-0.064**	-0.023	-0.025
	(0.032)	(0.033)	(0.033)	(0.032)	(0.033)	(0.033)
Female	0.082***	0.082***	0.082***	0.082***	0.083***	0.082***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
White	-0.243***	-0.248***	-0.229***	-0.241***	-0.247***	-0.227***
	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)
MP's Age		0.003***	0.003***		0.003***	0.003***
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.012***	-0.011***		-0.012***	-0.012***
		(0.003)	(0.003)		(0.003)	(0.003)
MP Single		-0.053	-0.053		-0.054*	-0.054*
		(0.033)	(0.033)		(0.032)	(0.032)
MP Divorced/ Other		0.019	0.021		0.019	0.022
		(0.029)	(0.029)		(0.029)	(0.029)
MP Missing		-0.051	-0.052		-0.052	-0.053
		(0.081)	(0.082)		(0.081)	(0.082)
SP Single		-0.173**	-0.170**		-0.173**	-0.170**
		(0.083)	(0.080)		(0.082)	(0.080)
SP Divorced/ Other		0.065	0.054		0.063	0.052
		(0.069)	(0.069)		(0.069)	(0.068)
SP Missing		-0.057*	-0.059*		-0.058*	-0.060*
		(0.031)	(0.031)		(0.031)	(0.031)
North East			-0.041			-0.044
			(0.030)			(0.031)
North West			-0.074***			-0.075***
			(0.016)			(0.016)
Yorkshire and Humber			-0.029*			-0.032*
			(0.017)			(0.017)
East Midlands			-0.056***			-0.054***
			(0.018)			(0.018)
West Midlands			-0.043***			-0.043***
			(0.014)			(0.014)
East of England			-0.058***			-0.059***
			(0.016)			(0.016)
Class size			-0.008***			-0.009***
			(0.003)			(0.003)
Constant	0.962***	0.806***	1.011***	0.966***	0.804***	1.025***
	(0.014)	(0.043)	(0.073)	(0.014)	(0.043)	(0.075)
Observations	7469	7469	7469	7469	7469	7469

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Table B.5.2: LPM and IV-2SLS estimates of peers' aspirations on age 16 aspiration to study for academic/vocational qualifications (Aspiration 5)

	LPM			IV-2SLS		
	Spec 1	Spec 2	Sepc 3	Spec 1	Spec 2	Spec 3
	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Aspiration 5	0.256***	0.243***	0.209***	0.143	0.072	0.038
	(0.033)	(0.034)	(0.036)	(0.097)	(0.102)	(0.112)
KS2 score	0.147***	0.143***	0.143***	0.150***	0.146***	0.146***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
MP A levels	-0.034**	-0.029**	-0.028**	-0.036***	-0.032**	-0.030**
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
MP Other Qualification	-0.068***	-0.063***	-0.063***	-0.072***	-0.069***	-0.068***
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)
MP No Qualification	-0.070***	-0.072***	-0.072***	-0.072***	-0.075***	-0.075***
	(0.014)	(0.014)	(0.015)	(0.014)	(0.015)	(0.015)
MP Missing	0.022	0.021	0.023	0.021	0.021	0.023
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
SP A levels	-0.025*	-0.021	-0.022	-0.027*	-0.023	-0.023
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
SP Other Qualification	-0.040***	-0.034**	-0.034**	-0.044***	-0.038***	-0.038***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
SP No Qualification	-0.036**	-0.029*	-0.030*	-0.039**	-0.033**	-0.034**
	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)
SP Missing	-0.011	-0.006	-0.005	-0.014	-0.010	-0.009
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
MP Self Employed	0.025	0.021	0.022	0.025	0.020	0.022
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
MP Unemployed	-0.004	0.007	0.010	-0.003	0.008	0.012
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
MP Out of the labour	0.007	0.017*	0.017*	0.007	0.017*	0.017*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
MP Missing	0.046	0.036	0.039	0.042	0.030	0.036
	(0.039)	(0.040)	(0.040)	(0.040)	(0.041)	(0.041)
SP Self Employed	-0.008	-0.009	-0.010	-0.008	-0.009	-0.010
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
SP Unemployed	-0.022	-0.018	-0.019	-0.019	-0.014	-0.017
	(0.033)	(0.033)	(0.033)	(0.034)	(0.033)	(0.034)
SP Out of the labour	0.015	0.018	0.018	0.015	0.018	0.019
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
SP Missing	-0.067**	-0.027	-0.028	-0.067**	-0.024	-0.025
	(0.031)	(0.032)	(0.032)	(0.031)	(0.033)	(0.033)
Female	0.079***	0.080***	0.080***	0.080***	0.082***	0.082***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
White	-0.211***	-0.218***	-0.212***	-0.224***	-0.238***	-0.225***
	(0.010)	(0.010)	(0.011)	(0.015)	(0.016)	(0.013)
MP's Age		0.003***	0.003***		0.003***	0.003***
		(0.001)	(0.001)		(0.001)	(0.001)
SP's Age		0.001	0.001		0.001	0.001
		(0.001)	(0.001)		(0.001)	(0.001)
Number of siblings		-0.012***	-0.011***		-0.012***	-0.012***
		(0.003)	(0.003)		(0.003)	(0.003)
MP Single		-0.058*	-0.056*		-0.056*	-0.054*
		(0.033)	(0.033)		(0.033)	(0.033)
MP Divorced/ Other		0.014	0.016		0.018	0.021
		(0.029)	(0.029)		(0.029)	(0.029)
MP Missing		-0.059	-0.057		-0.055	-0.054
		(0.080)	(0.081)		(0.081)	(0.082)
SP Single		-0.165**	-0.165**		-0.171**	-0.169**
		(0.083)	(0.082)		(0.082)	(0.080)
SP Divorced/ Other		0.064	0.058		0.063	0.053
		(0.068)	(0.068)		(0.068)	(0.068)
SP Missing		-0.051*	-0.053*		-0.056*	-0.058*
		(0.031)	(0.031)		(0.031)	(0.031)
North East			-0.013			-0.038
			(0.025)			(0.035)
North West			-0.044***			-0.069***
			(0.014)			(0.023)
Yorkshire and Humber			-0.007			-0.028
			(0.015)			(0.021)
East Midlands			-0.031**			-0.050**
			(0.015)			(0.021)
West Midlands			-0.020			-0.039**
			(0.013)			(0.018)
East of England			-0.031**			-0.054**
			(0.014)			(0.021)
Class size			-0.005**			-0.008**
			(0.002)			(0.003)
Constant	0.746***	0.615***	0.776***	0.847***	0.748***	0.980***
	(0.031)	(0.048)	(0.072)	(0.085)	(0.091)	(0.155)
Observations	7469	7469	7469	7469	7469	7469

Robust standard errors clustered at the secondary school level reported in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Omitted groups: Degree; Employed; London; Married.

Chapter 4

Socio-economic background and academic match of students in 16-19 post-compulsory education

4.1 Introduction

4.1.1 Background and motivation

Improving the quality of the educational system has been an important concern for policy makers as it has been proven to be an effective route away from poverty and a key path to economic prosperity and social inclusion (Black et al., 2015, Jenkins et al., 2003). In light of this, various investigations attempted to raise awareness of the factors causing inequality and social immobility in education. However, little is yet known about the extent to which the type of education the students are exposed to matches their ability credentials and most importantly whether certain groups of students, especially those who come from a disadvantaged background, are more likely to be exposed to academic mismatch. Several studies reveal that students from low-income families or students from ethnic minorities tend not to attend a college or to disproportionately attend less selective colleges, drop out before attaining a degree or graduate with a lower grade (Hearn, 1991, Hoxby and Avery, 2012, Chowdry et al., 2013, Crawford, 2014, Crawford et al., 2016). These poor college outcomes are often attributed to disadvantaged students being less academically prepared for college than their more advantaged counterparts (Hoxby and Avery, 2012).

In response to these concerns there is an emerging literature which investigates Higher Education (HE) institution quality and ability match of individuals from disadvantaged backgrounds. The present study aims to contribute to this literature through investigating the relationship between socio-economic background and academic match at the upper secondary level. I am able to do this through making use of very detailed administrative records from schools and tax authorities in England. Post-compulsory (non-tertiary) education usually lasts for two years and it is received between the ages of 16-18 in the

form of upper secondary (Level 3) qualifications[21], although it is very common for the students to remain in post-compulsory (non-tertiary) education until age 19. Also, the students in upper secondary education have the option to select between an academic or a vocational course which usually take place in different types of institutions. For example, academic upper secondary qualifications are usually taught in sixth form colleges or sixth form schools while vocational upper secondary qualifications are often studied in Further Education colleges. This study is the first in the field of course-ability match, at least to my knowledge, to consider both academic and vocational qualifications.

In 16-19 post-compulsory education, academic undermatch would occur when a student's achievements at the final year of compulsory school, at age 16, would permit access to more highly ranked qualifications than the ones they actually choose. On the other hand, academic overmatch is when students choose to study for more highly ranked qualifications when their prior academic achievements are below those typically seen at that level. Upper secondary qualifications are of major importance for young people's progression as these qualifications determine their entrance into the labour market or their enrolment in HE.

One could, of course, argue that the students when choosing post-16 education are behaving optimally. Theoretically, non-matched students are not necessarily less likely to do well in their chosen qualification and they might enjoy all other dimensions of a qualification beyond its academic selectivity. For example, we might define a student as being undermatched because of having the skills to study A levels in three facilitating subjects[22] but did not. The student, though, might have enjoyed studying the non-facilitating subject for example in Arts or Music. Moreover, undermatched students might gain utility from being the 'big fish in a small pond' (Marsh and Hau, 2003) or from the fact that they will have to deal with less competition from their peers and be able to manage coursework better (Campbell et al., 2019, Dillon and Smith, 2017). On the other hand, overmatched students might benefit from having stronger peers which could lead to better attainment.

However, there are several possible reasons why being undermatched or overmatched might not be desirable. First, it is supported that *academic assortative matching*[23] maximises

[21]Post-compulsory (non-tertiary) education that is referred to in the study is the equivalent education received in the final two years of senior high school in the USA.

[22]Facilitating subjects are the subjects most commonly required or preferred by universities to get accepted onto degree courses. Many of the top universities require students to have at least one A level in a facilitating subject when they apply.

[23]Academic assortative matching involves the preferential matching of students to qualifications studied by similarly achieving peers. In other words, academic assortative matching would occur when the most high-achieving students study for the most highly ranked qualifications.

the efficiency of human capital production (Sallee et al., 2008). In addition, students of all academic ability levels have a higher probability of completing a course if the selectivity level matches their measured academic level (Hoxby and Avery, 2012). Also, undermatched students may have lower quality peers and may therefore be negatively influenced. Further, if the individuals would like to pursue HE then upper secondary qualifications will play an important role on the probability of being accepted on each course and in each university. Finally, there are employers, for example employers who hire undergraduate students for advanced apprenticeships, who value upper secondary qualifications a lot. Academic match in 16-19 post-compulsory education, therefore, has important implications for social mobility and the life chances of students. Nevertheless, I make no claims that obeying academic assortative matching in upper secondary education is individually optimal for *all* students.

4.1.2 Research question

The chapter examines the relationship between socio-economic background and academic match among students in post-compulsory (non-tertiary) education using a cohort of young individuals born in 1989-1990[24]. Making use of a unique standardised matching index which identifies undermatched, matched and overmatched individuals in the upper secondary level, the study aims to uncover socio-economic inequalities causing academic mismatch. In addition, the study aims to take this analysis a step further and investigate additional possible correlations between students' background and academic match. It explores regional differences between the most undermatched and the most overmatched students as well as differences in their school environments by comparing school composition characteristics such as peers' ability and socio-economic background. Finally, the study examines the short-term labour market returns of academic assortative matching on income earned at age 25.

4.1.3 Research findings and limitations

The main findings of this study are summarised as follows:

[24]As explained in Chapter 1, prior to recent policy allocations the students could leave education and directly enter the labour market at age 16 without completing any upper secondary course. Since 2016 the compulsory participation age rose to 18 but the students still have to select between an academic or a vocational route. The cohort used in the analysis turned 16 before the policy change and therefore they were not affected. This study considers only the students who continued in post-compulsory upper secondary education which is about 80% of the students in the total cohort.

- Disadvantaged students are more likely to be exposed to academic undermatch even when compared to students from the same school.
- Academically undermatched students are more likely to be found in disadvantaged schools including schools with lower proportions of high achieving students and higher proportions of students who are eligible for a Free School Meal (FSM).
- There are indications that critical masses of undermatched students are more likely to be found in rural Local Authorities Districts (LADs)[25] with higher rates of youth unemployment and higher proportions of poorly educated residents.
- Academic assortative matching in 16-19 post-compulsory education has a positive relationship with labour market income returns, at least at early ages.

Ideally this study which examines the impact of socio-economic background on academic match would require the data to include rich individual-level information on students' socio-economic background. Unfortunately, the administrative data used, although very rich in information about individuals' prior achievements and post-16 qualifications, are weak in respect to socio-economic characteristics. I overcome this limitation by combining both individual and detailed neighbourhood level measures to identify each student's socio-economic position as will be explained thoroughly in Section 4.3. Chowdry et al. (2013), who used in their study the same data and a similar approach to identify socio-economic background, checked the validity of this approach using the Longitudinal Study of Young People in England (LSYPE), which uses a sample of the same cohort and includes detailed socio-economic indicators. Their analysis showed that the index of socio-economic status created using the administrative data is successful in ranking pupils according to individual measures of socio-economic status, including household income, mother's education, father's occupational status and housing tenure.

4.1.4 Structure

The remainder of this chapter is organised as follows. In the next section there is a review of the limited current literature on academic match and a discussion of the contribution of this study. Section 4.3 describes the data and provides a deliberation on academic match and local area characteristics. Section 4.4 explains the methodology that has been used to

[25] A Local Authority District (LAD) is a term used to describe geographic areas in England including London Boroughs, Metropolitan Districts, Unitary Authorities and Non-Metropolitan Districts.

examine the relationship between socio-economic background and academic match and Section 4.5 presents the main findings. Finally, Section 4.6 concludes this chapter.

4.2 Literature Review

4.2.1 Existing literature on academic match

There is a relatively small but emerging literature in economics that examines the phenomenon of academic match and its implications, mainly on university performance and institution prestige. In the current literature, academic undermatch occurs when a student's academic capacity would allow access to a college or university that is more prestigious than the alternative institution they choose. Hoxby and Avery (2012) find that the vast majority of high-attaining students who come from a disadvantaged background do not apply to selective colleges or universities despite the fact that selective institutions would often cost them less due to generous financial aid available for disadvantaged students. The authors also note that the disadvantaged students who do not apply to any prestigious institution despite their high academic achievements tend to come from areas too small to support selective public high schools. Also, compared to other high-achieving disadvantaged students who do apply to prestigious institutions, they tend to have lower achieving peers and are unlikely to encounter a teacher or a schoolmate from an older cohort who attended a selective institution.

Dillon and Smith (2017) defined student-college academic mismatch as the difference between a student's percentile position in the cognitive ability distribution and the student's college percentile position in the college's quality distribution. Using a sample from the USA born in 1987 the authors identified substantial amounts of both academic undermatch and academic overmatch from students of all socio-economic backgrounds. The study mentions, though, that students from less wealthy families undermatch more while more informed students, such as students from high schools where many graduates go to college, undermatch less and overmatch more. Further, the study identifies that academic mismatch is driven mostly from student application and enrolment decisions rather than college admission decisions.

Smith et al. (2013) followed two cohorts of American students, the older finishing school in 1992 and the more recent in 2004. They determined the highest academic selectivity

college to which the students had access given their academic credentials and compared that college to where the students ultimately enrolled by making use of data on students' applications and admission offers. The study identified that over 40% of the students undermatched and that undermatch is more common among students from rural areas and for students from low socio-economic backgrounds with less educated parents. Finally their findings show that undermatch has decreased between the two cohorts partly because of changing student decisions and because of changing college selectivity.

The study of Roderick et al. (2011) showed that low income urban American students who had the qualifications to attend four-year colleges do not effectively take the steps to apply to and enrol in a four-year college. Also, they found that students enrol in colleges with selectivity levels below the ones that they would be qualified to attend based on their achievements. The study showed that high schools have an important role to play in guiding students into the college application process and shaping the students' choices. Students who attended schools with a higher proportion of college-attending students, where teachers report that they expect students to go to college and take responsibility for their students' college application, and where greater proportions of students are active in financial aid application are more likely to plan to attend, apply to, and be accepted into a college as well as to enrol in a college with selectivity levels that match their qualifications.

Belasco and Trivette (2015) using the same data but a different strategy to identify academic match have also shown that there is a negative and significant relationship between socio-economic background and academic match especially for students living in rural areas. However, the study revealed that the influence of other determinants such as school environment, the provision of information and other college-related interventions play a more important role for academic match than a student's socio-economic background.

Black et al. (2015) focused their study on academic undermatch of high-achieving students from an ethnic minority background using data collected by the Texas Workforce Data Quality Initiative. The study defined undermatch as not applying to a top-tier flagship university while graduating from high school in the top 10% or in the top 25% of the senior class's students during 2008 and 2009. The study identified that the phenomenon of undermatch exists even if students have perfect information that they will be admitted. Also the authors mention that automatic university admission contributes to minority applications to elite universities but is not sufficient to fully overcome academic undermatch.

All of the literature discussed above is focused on the USA. The closest paper to the present study is that of Campbell et al. (2019) which is the only study to my knowledge that examines academic undermatch for HE students in England. The study creates a new continuous measure of student-course match quality representing the distance of each student's university course from their ideally matched course, which would be that attended by others in the same position in the ability distribution. The study of Campbell et al. (2019) uses the same detailed administrative records as the ones used in this study and for the same cohort of students born in 1989-1990. The authors identified that high-attaining disadvantaged students are more likely to be undermatched at their university course than their more advantaged counterparts and also that undermatched students are indicated to live in environments which are less conducive to academic success such as areas that have fewer residents who attended university or in areas where there are fewer universities.

4.2.2 Contribution to the literature

In the reviewed literature there is a general consensus about the role of socio-economic background on students' academic match in HE. Students from disadvantaged backgrounds are more likely to be exposed to academic undermatch in their chosen university course. This study builds on the methods of Campbell et al. (2019) by creating a new definition of academic match which identifies undermatched, matched and overmatched students in 16-19 post-compulsory education.

The study is unique for three main reasons. First, it is the first study to my knowledge to study academic match at the upper secondary level and it is also the first study to consider students' match in vocational qualifications, as all the previous studies considered only academic education as a possible option. Also, it is only the second study to examine academic match in England in general, after Campbell et al. (2019), as most of the previous studies are focused on the USA. In opposition to the USA where the final two years of senior high school are compulsory for all students, in England, until very recently, compulsory schooling ended at age 16. Although the vast majority of students remained in education the options available concerning where to go and what to study were, and still are, very broad. The students have the option to go to a sixth form in their school (if there is a sixth form in their school) or in another school and study for academic qualifications, they can also go to a sixth form college and finally, they can go to a college of Further

Education were they can study mostly vocational qualifications. Also, while for the students undertaking A levels (which are the main academic qualifications offered) the route is relatively well-known, for other students who study mainly vocational qualifications, the available options are much more diverse and not as easy to understand where they lead to (Hupkau et al., 2017). Given, therefore, the complexity of the vocational education system in England (Wolf, 2011, Hupkau et al., 2017), the broad list of options the students have to choose from and the fact that about 50% of the students staying in post-compulsory education do not undertake A levels and are therefore exposed to this complex vocational system, an investigation as to whether the course they finally end up studying matches their ability credentials should be of major interest to policy-makers.

Second, the study takes this investigation on academic match at the upper secondary level a step ahead by exploring regional differences between the most undermatched and the most overmatched students as well as differences in their school environments by comparing school composition characteristics such as peers' ability and socio-economic background. Finally, the study examines the short-term labour market returns to academic assortative matching, using measures of the students' income at age 25, attempting to shed light on the importance of being academically assortatively matched.

4.3 Data

4.3.1 Description of the Datasets

The analysis of this study has been carried out by making use of individual-level linked administrative data on the whole population of state-school students in England for a single cohort totalling approximately half a million students. This cohort was born in 1989-1990, took compulsory age 16 public examinations (GCSE exams[26]) in 2006 and attended 16-19 post-compulsory education between 2006-2009. Table 4.1 below outlines the educational progression of the cohort.

The school data come from the National Pupil Database (NPD) and include basic demographic characteristics as well as school outcomes at age 11 and age 14 (KS2 and KS3), public examination results at age 16 (KS4: GCSEs) and age 18 (KS5: usually A levels). The data for vocational qualifications come from the Individualised Learner Record

[26]GCSEs are compulsory high-stake public examinations taken by all school students usually in 9 or 10 subjects at the end of lower secondary education at age 16.

Table 4.1: Education progression of the cohort

Academic Year	Age	Event
1989-1990	0	Born
2000-2001	10-11	Key Stage (KS) 2 exams
2003-2004	13-14	KS3 exams
2005-2006	15-16	KS4 exams (GCSEs)
2007-2008	17-18	KS5 exams (upper secondary education)
2008-2009	18-19	KS5 exams for continuing students

(ILR) dataset which is an administrative dataset covering the population of funded learners in Further Education in England. The ILR contains detailed information on the learning undertaken by individuals, including the learning aim, type of qualification, level, subject area, training provider, start and end dates, and attainment markers. The aggregated earnings data come from the Longitudinal Education Outcomes (LEO) dataset which are compiled from HMRC tax records by officials at the Department for Education in the UK.

The data have been also linked to the School Census of 2006 in order to access important information about the composition of schools the cohort attended. Finally, the datasets have been linked to the 2001 Census in order to acquire detailed measures of neighbourhood composition and Local Authority District (LAD) characteristics.

4.3.2 Analytical Sample

The analytical sample is composed of students who went to state-schools in England and achieved at least one post-compulsory upper-secondary qualification. From the initial cohort of about 555,000 students there is a moderate censoring since a proportion of the students left education after compulsory examinations at age 16 or attempted an upper-secondary qualification but did not achieve it (around 23% of the cohort)[27]. The final analytical sample is composed of over 390,000 students with non-missing information on socio-economic background, test scores at ages 11 and 14 and exam scores at age 16, demographics at age 16 and secondary school attended. To construct variables such as the socio-economic background index, GCSE scores to identify the selectivity of each upper-secondary qualification and to standardise prior achievement (KS2, KS3 and KS4) the total sample for whom information was available has been used. Table 4.1 shows the attrition details of the sample while in Appendix C.1 it is provided a detailed comparison of

[27]In the future it would be interesting to consider academic match in 'highest attempted' upper-secondary qualification rather than 'highest achieved' qualification and examine the extent to which students who where matched to their qualification had a higher probability of achieving it.

the total and analytical sample. The analytical sample has a higher average prior attainment than the total sample which is not surprising given that the study is focused only on post-compulsory education participants who achieved an upper-secondary qualification. We would expect that these students would be higher achieving than the total sample which includes the students who dropped out from education or had attempted a post-compulsory qualification but did not achieve it. Apart from that, in terms of other background characteristics, the analytical sample remains representative of the total sample.

Table 4.1: Attrition Details

Sub-sample	Observations
Total in cohort	555,601
Total post-compulsory education participants	428,802
Non-missing background characteristics	**391,651**

4.3.3 Variables

Dependent Variable

Step 1: Ranking post-16 qualifications in order of selectivity

The analysis is based on categorising the many different types of post-16 qualifications available into five broad categories. These categories include Level 3 academic qualifications, including A levels and AS levels, Level 3 vocational qualifications, Level 2 vocational qualifications and Level 1 vocational qualifications. Making use of information published by the Russell-Group (selective) university guidance[28] I am able to distinguish between students who studied for A levels which include three or more facilitating subjects, two facilitating subjects, one facilitating subject and those who studied for A levels but not in a facilitating subject[29]. The outcome is a categorical variable of 11 different classifications indicating the highest upper secondary qualification each individual achieved by age 20[30].

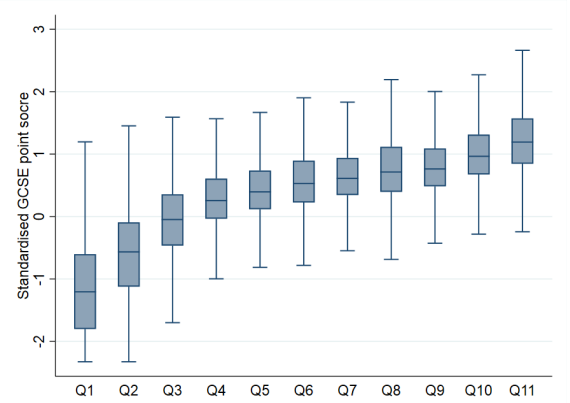
[28]More information about the Russell Group guidance can be found by accessing their web-page at https://russellgroup.ac.uk.

[29]According to the Russell-Group university guidance the A level subjects which are most often required, called "facilitating subjects" include the following: Mathematics and Further Mathematics, English Literature, Physics, Biology, Chemistry, Geography, History and Languages. I used the taxonomy provided by Dilnot (2018) to distinguish between two groups: facilitating and non-facilitating A-levels. Dilnot (2016) suggested that one of the reasons why students from lower socio-economic backgrounds are under-represented at high status universities is because of differential choice of A levels. These differentials, though, substantially disappeared when the author accounted for age 16 test scores. In any case, A levels in facilitating subjects are more valued by institutions and it is considered important to distinguish between the facilitating and the non-facilitating subjects.

[30]There are many students who at 16 do not meet the pre-requisites of Level 3 qualifications and must study at Level 2 (i.e. at the same level as GCSE) or even Level 1 or below that (at Entry Level). Although it is possible for students to repeat the GCSE examinations instead of pursuing a non-compulsory vocational qualification in this study I consider only the students studying for Level 1 and Level 2 vocational

Then, I identified the selectivity of each qualification based on the median standardised GCSE point score[31] of the students studying for that qualification. Figure 4.1 shows the distribution of the scores of students in each qualification in order of selectivity[32].

Figure 4.1: Measure of qualification's selectivity based on median standardised GCSE scores of students studying for that qualification



Q1: Level 1 Vocational Qualifications
Q2: Level 2 Vocational Qualifications
Q3: Level 3 Vocational Qualifications
Q4: AS Levels
Q5: A levels<3 without facilitating subjects
Q6: A levels<3 with 1 facilitating subject
Q7: A levels≥3 without facilitating subjects
Q8: 2 A levels in 2 facilitating subjects
Q9: A levels≥3 with 1 facilitating subject
Q10: A levels≥3 with 2 facilitating subjects
Q11: A levels≥3 with facilitating subjects≥3

Note: The box plots display the full range of variation of GCSE scores of students achieving each qualification with the upper and lower lines of the box representing the 75^{th} and 25^{th} percentiles of GCSE scores respectively, the middle line representing the median GCSE scores and the top and bottom extending lines the range.

It is evident that there are substantial differences in the 'selectivity' of each qualification with the median scores of students studying for the most highly ranked qualifications being considerably higher than those of the students studying for the lowest ranked ones. In addition, the difference between the median GCSE scores of students studying for vocational and academic qualifications is also notable. We would expect to see this difference for Level 1 and Level 2 vocational qualifications, given that the students pursuing these qualifications are usually those who have not done well enough in their GCSEs and presumably could not pursue a Level 3 qualification. On the other hand, especially given that Hupkau et al. (2017) found that A levels and vocational qualifications at Level 3 are equally strong predictors of staying on in education up to the age of 18 and achieving a Level 3 qualification before the age 20, probably we should not expect this difference in the median test scores of students pursuing Level 3 vocational qualifications from those pursuing Level 3 academic qualifications.

qualifications. An added complication in identifying student's post-16 educational route is that there are students who are engaged in several different qualifications at different levels and of different types simultaneously. The different levels of qualifications is not of significant importance in this study since I consider only the highest qualification achieved. Considering the different types of qualifications, there is a very small proportion of students (less than 1% of the sample) who achieved both a Level 3 academic and a Level 3 vocational qualification. In that case, the students have been categorised as studying for A levels since these qualifications are usually the dominant route to university and also it is rare that people with other types of qualifications make it to a Russell Group university (Hupkau et al., 2017).

[31]The GCSE point score is calculated from the total number of GCSE points of the students' best eight subjects including Maths and English.

[32]The qualifications have been implicitly ranked in ascending order representing the selectivity of each, based on the median GCSE point score of the students studying for each qualification.

Step 2: Creation of the Academic Matching Index

Following the method introduced by Campbell et al. (2019), a new continuous measure of student-qualification match has been produced which identifies academic match in 16-19 post-compulsory education. The matching index is calculated using the following steps:

1. **Individual ability distribution:** I identify the ability of each student based on their position in the standardised distribution of the GCSE point score.

2. **Qualification quality distribution:** I identify the quality (or selectivity) of each group of upper secondary qualifications (Q1-Q11) in a distribution of qualification quality based on the standardised median GCSE point score of students who achieved that qualification.

3. **Academic matching index:** I subtract the student's position on the individual ability distribution from the position of their course in the qualification quality distribution. In other words, the academic matching index *subtracts* the standardised GCSE point score of the individual from the median GCSE point score of the students who achieved the same qualification as the individual. In mathematical terms the equation used is:

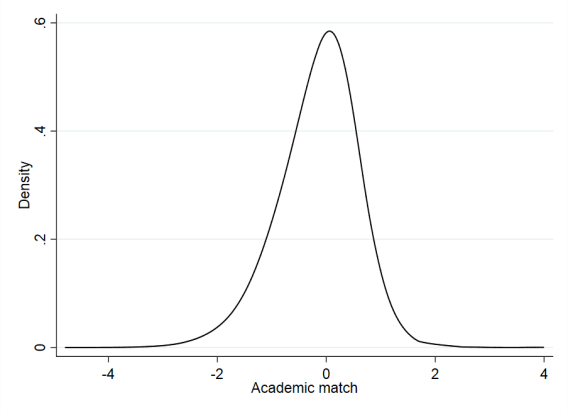
$$Match_{i,q} = med(zgcse_q) - zgcse_i$$

.

The result is a continuous measure of academic match for each student in post-compulsory education. Figure 4.2 shows the distribution of the measure of academic match resulting from the process explained above. The matching index presents the distance of each student's standardised GCSE point score from the median standardised GCSE point score of the students' achieving his or her chosen qualification. This method benefits from being able to identify low and high levels of undermatch or overmatch. The students are undermatched or overmatched on a standardised scale ranging from -4.66 to 3.52. A negative value on the matching index implies that the student is studying for a qualification that is lower on the qualification quality distribution than they are on the individual ability distribution (in other words, the student is undermatched). A positive value implies that the student is studying for a qualification that is higher on the qualification quality distribution than they are (in other words, the student is overmatched). Finally, for students who scored close to zero on the matching index it is indicated that their qualification quality is matching their ability level. If a student is exactly average on the individual ability

distribution then, the student is matched if their chosen qualification is exactly average in the qualification's quality distribution. If the student is one standard deviation above the average in the ability distribution and they choose an upper secondary qualification which is exactly average in the quality distribution, then they are undermatched by one standard deviation[33].

Figure 4.2: Measure of student-qualification match



The basis of this continuous points-based measure of academic match is the idea that a student should be broadly comparable (matched) with their equally able peers in terms of educational attainment (Campbell et al., 2019). This idea has been supported in all studies on academic match in HE reviewed in Section 4.2 and a similar approach has been attempted to be applied in this study for upper secondary qualifications. Especially for upper secondary qualifications, though, it should be recognised that such definitions of qualification quality are somewhat subjective particularly given that the scope of A levels and AS levels differs from that of vocational qualifications. The age 16 points-based indicator of qualification rank is not necessarily the most important or the only important factor in determining the quality of the qualification.

However, obtaining A levels, especially in facilitating subjects, provides a greater opportunity to enter a high status HE institution which is associated with a higher wage return (Chowdry et al., 2013). It is also suggested that vocational education prevents drop-out rates (De Groote, 2017), presumably of the least able students who are likely to be more vulnerable to drop-outs as well as reducing unemployment rates of young people. I will therefore argue that the points-based measure of a qualification's rank is a good proxy for the nature of the upper post-secondary qualification being studied by a particular student

[33]As will be explained later on, when having to categorise students as being undermatched, matched and overmatched for some of the analyses that follow, I relaxed the cut-off point and considered students as being matched if they scored between half a standard deviation above or below zero on the standardised matching index.

which in turn will have long run economic implications for these individuals.

Key Variable

Socio-economic background

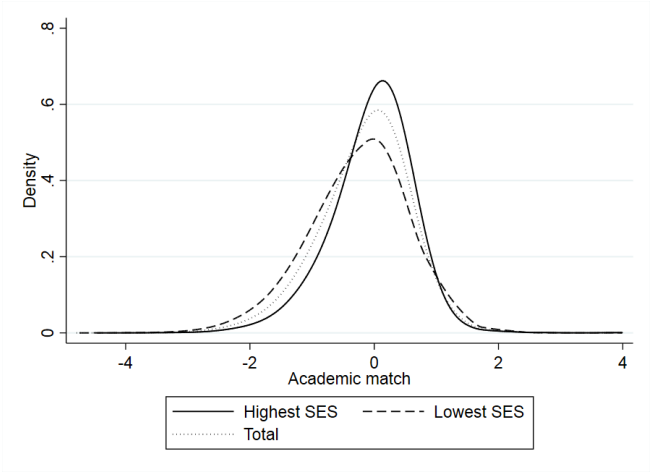
The key variable in this study is the socio-economic group the student belongs to. Ideally for this study I would want a detailed measure of the student's socio-economic background such as family income, parental occupation and education. Unfortunately, the administrative data used in this study do not contain such measures of socio-economic background. Instead, I constructed a socio-economic index which combines both individual and neighbourhood level measures of socio-economic background. The individual measure of socio-economic background is collected from information concerning whether the student was eligible for Free School Meal (FSM) at age 16 which is basically an indicator of whether the student is from a household which receives state benefits. Given only a small minority of households are eligible for FSM, this is therefore not a sufficient measure for all students. Also, I need to be able to identify students who are from the highest and the lowest socio-economic groups and the FSM indicator would not do so. For that reason I make use of a set of neighbourhood variables taken from the 2001 Census and the 2007 Index of Multiple Deprivation (IMD). These measures are available at the 'Lower Super Output Area (LSOA)' level which is a geographic neighbourhood comprising of around 700 households or around 1,500 individuals. The approach I follow is very similar to that of Chowdry et al. (2013) and Campbell et al. (2019). The socio-economic index consists of:

1. Free School Meal eligibility (2006)
2. Index of Multiple Deprivation (2007)
3. Three 2001 Census local area-based measures indicating the proportion of neighbours that:
   (a) Work in managerial and professional occupation
   (b) Hold a Level 3 qualification or above
   (c) Are home owners

I combined these measures using *principal components analysis (PCA)* to create a standardised index using the whole population of state-school students at age 16 including those who did not participate in 16-19 post-compulsory education. Therefore, throughout this

chapter socio-economic background refers to the socio-economic position of a student relative to the whole school-cohort population for whom information was available. I then divided the sample into five socio-economic quintiles on the basis of this index. Figure 4.3 shows the distribution of academic match of students in the highest and lowest socio-economic quintile compared to the distribution of academic match of the total analytical sample. It is illustrated that students from the lowest socio-economic quintile are less well matched to their chosen post-secondary qualification compared to students from the highest socio-economic quintile.

Figure 4.3: Academic match of highest and lowest SES students



Control Variables

Measures of prior attainment

The measures of prior attainment come from school tests taken at the age of 11 and teacher assessments at the age of 14 reflecting the students' attainment in three core subjects, Maths, Science and English. The average age 11 test scores (average attainment in KS2 Maths, Science and English) and average age 14 level scores (average level achieved in KS3 Maths, Science and English) are used as indicators of prior attainment.

In addition to controlling for prior attainment at age 11 and 14, the students' performance in GCSE exams (in about 10 subjects) is expected to have a significant impact on students' academic match as well as a strong correlation with important variables such as socio-economic background. Given how the matching index is constructed, including the individual GCSE point score in the set of prior attainment controls has both its benefits and limitations. Regressing the matching index ($med(zgcse_q) - zgcse_i$) against individual score ($zgcse_i$) will show the impact of socio-economic background (and other covariates)

on upper-secondary qualification rank, conditional on individual GCSE point score, rather than the effect on match distance. On the other hand, not controlling for individual GCSE point score would produce biased results on any variable strongly correlated with it, including socio-economic background.

Figure 4.4 shows the distribution of academic match of students in the highest and lowest socio-economic quintile as well as that of the total sample across the individual GCSE point score distribution. There are two important points to be derived from Figure 4.4. First, the relative position of the two distributions (Highest SES and Lowest SES) reflects what has been also observed in Figure 4.3, that students from the lowest socio-economic quintile are less highly matched to their upper-secondary qualification across the GCSE point score distribution compared to students from the highest socio-economic quintile.

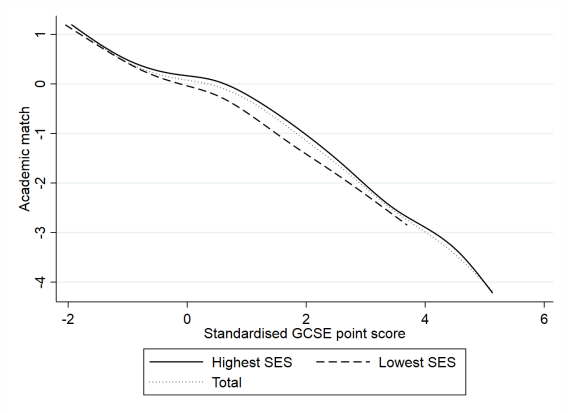
The second major point illustrated in Figure 4.4 is that academic undermatch is more prevalent among the high achieving students. This is not so surprising as low achieving students, with few GCSE points, could not undermatch since there would be no course whose median score would be at a lower position in the distribution. Therefore, we should expect that the socio-economic effect on academic match will vary significantly across students of differing ability levels and we can expect that the gap will be greater among high achieving students.

In order to address what is discussed above the estimations that follow combine both regressions with individual GCSE point score included in the set of prior attainment controls as well as heterogeneous analyses, splitting the sample into ability quintiles. In the latter, individual GCSE point score is not controlled for. All exam and test scores have been standardised to have a mean of 0 and a standard deviation of 1. The standardisation of these variables reflects the ability position of post-compulsory education participants compared to the total cohort of students.

Other individual and school characteristics

I control for demographics including gender and ethnicity recorded at age 16. Most highly ranked 16-19 courses require 5 A*-C GCSEs including Maths and English and therefore I also include an indicator of whether the student achieved this threshold in the set of demographic controls. I also attempt to control for school quality, peer effects and unobserved differences between students by applying school fixed effects taking into account the secondary school attended. It should be mentioned that although the school an

Figure 4.4: Academic match across the GCSE point score distribution



individual attends is likely to be an important determinant of academic match we should be very cautious about the interpretation of the results in the presence of these fixed effects for reasons that will be explained in Section 4.4.

4.3.4 Descriptive Statistics

The descriptive statistics of the analytical sample by quintile of socio-economic group are provided in Appendix C.2. The descriptive statistics indicate, first, that the average prior achievements of the analytical sample are higher than the average prior achievements of the total cohort (observed from the standardised test scores that have mean score above the average). As explained above, this is not surprising given that the study is focused only on students who continued in post-compulsory education who we would expect to be more high-achieving than the total cohort which includes students who dropped out from education.

Further, it is indicated that there is a significant difference by socio-economic group in the proportion of students leaving compulsory schooling having achieved 5 A*-C GCSEs including Maths and English with 69% of the students from the highest socio-economic group having achieved this threshold compared to just 29% of students from the lowest socio-economic group. Also, there are substantial differences in the average test scores of students from differing socio-economic groups with the achievements of the poorest students being significantly lower than those of the students from the highest socio-economic group both at school examinations at ages 11 and 14 and high-stake national examinations at age 16. It is also worth noting that students from ethnic minorities tend to be over-represented in the lowest socio-economic groups. Further, it can be observed that there are substantial

differences in the post-compulsory qualifications the students from different socio-economic groups achieved. It is evident that the most highly ranked qualifications are much more prominent for students in the highest socio-economic group as 39% of the students achieved the three most highly ranked qualifications, compared to only 10% of the students from the lowest socio-economic group.

4.3.5 Where are the undermatched and overmatched students in England?

The chloropleth map in Figure 4.5 shows the proportion of undermatched and overmatched students in each LAD of England. The darker is the LAD's colouring the higher the proportion of undermatched students living in it. The map demonstrates that there are several critical masses of undermatched students in mostly rural areas in Durham, Leicestershire and in the South West (see Figure 4.6 for information on rural and urban areas). On the other hand, critical masses of overmatched students can be observed in non-rural areas in Greater London and the South East including Surrey and Oxfordshire. The other critical masses of overmatched students are more scattered but some areas in South Cumbria and North Yorkshire can be picked out.

The incentives of students to acquire qualifications might partly be due to local labour market conditions or due to highly or poorly educated parents being concentrated in specific areas. To enlighten on this matter, the chloropleth map in Figure 4.7 illustrates the proportion of each district's residents who hold at least a Level 4 Qualification (equivalent to a Certificate of HE). Similarly, the chloropleth map in Figure 4.8 illustrates the proportion of residents aged 16-24 who claim Job Seekers Allowance (in other words the rate of youth unemployment). We can observe that there are some critical masses of highly educated residents and low proportions of claimants in Greater London, the South East and in South Cumbria and North Yorkshire. In districts such as Durham, Leicestershire and Cornwall, where critical masses of undermatched students have been observed, a high unemployment rate and a poorly educated population can be also observed.

Figure 4.5: Proportion of academic match in each LAD

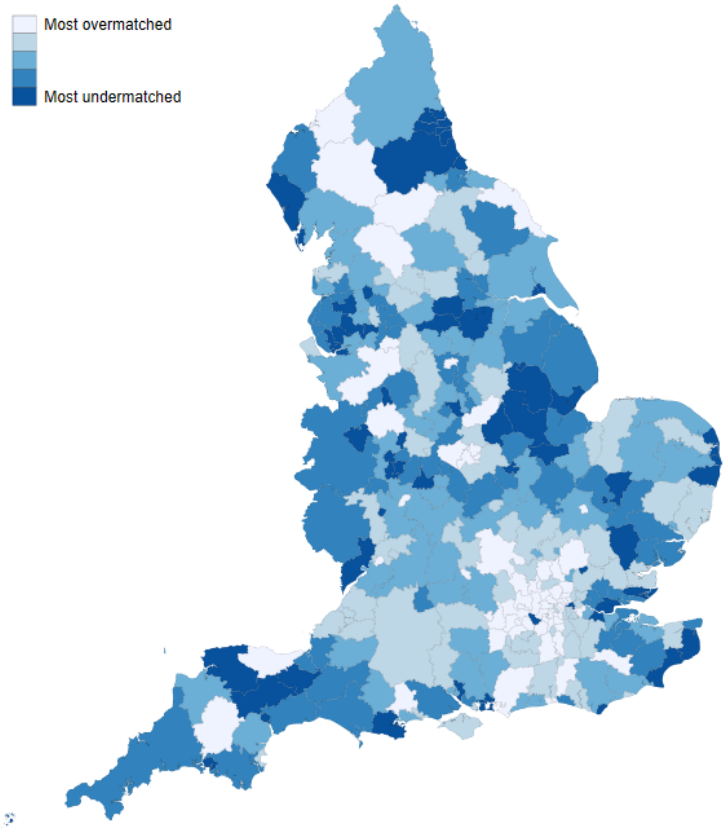


Figure 4.6: Urban and Rural LADs

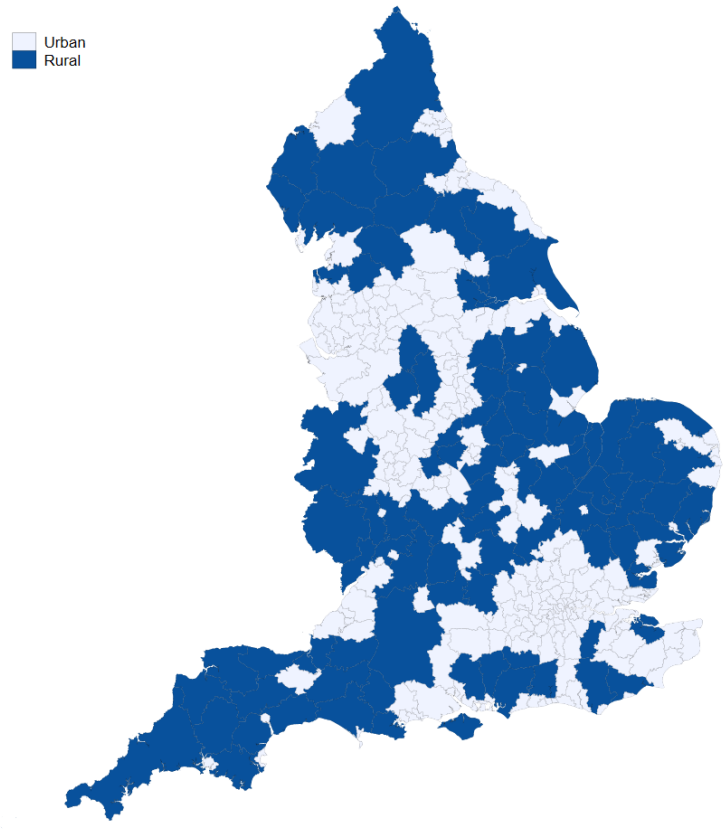


Figure 4.7: Proportion of population with a Level 4 Qualification or above in each LAD

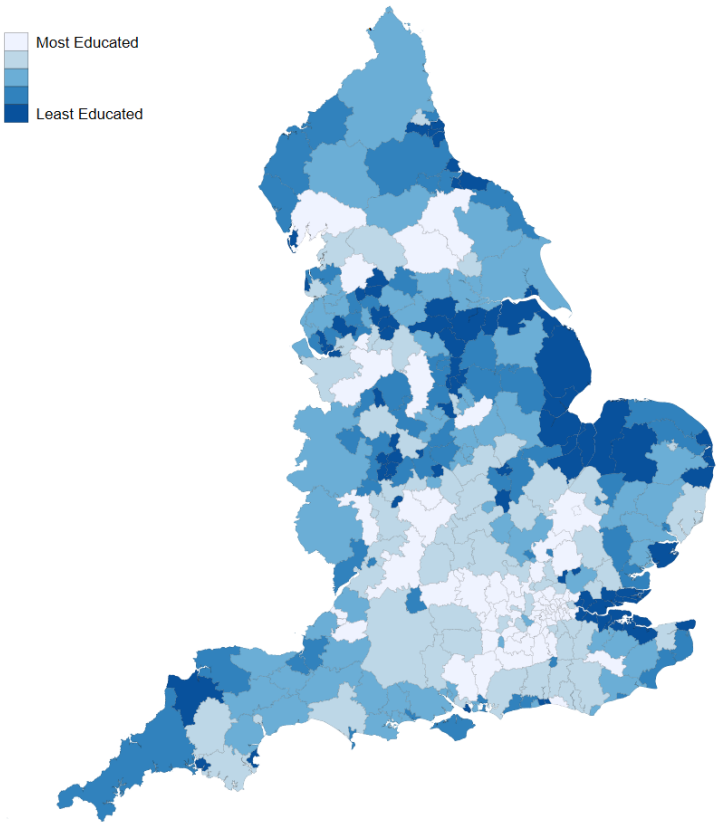
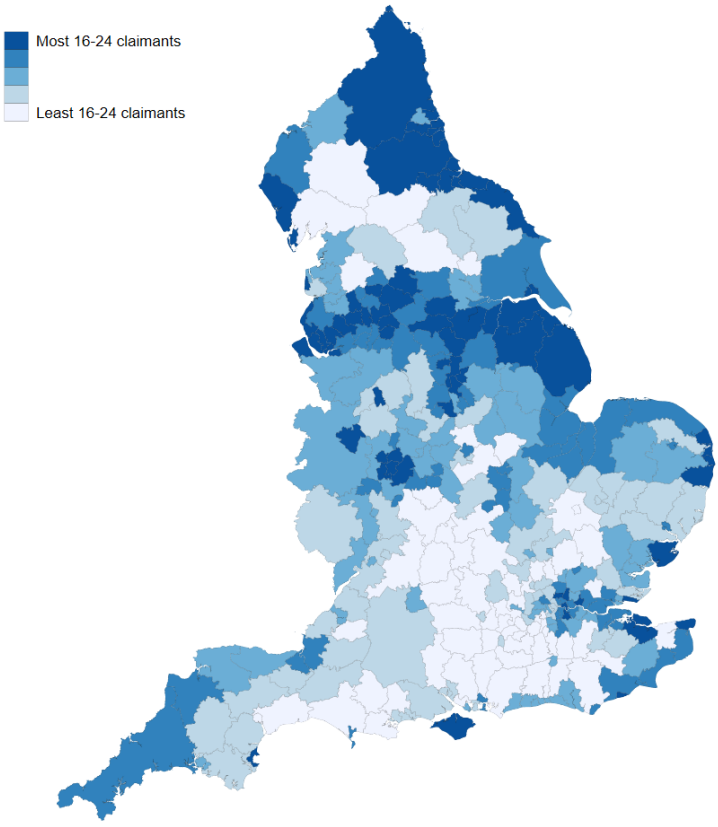


Figure 4.8: Proportion of 16-24 claimants in each LAD



4.4 Methodology

I estimate the following model for the i^{th} individual who participated in a 16-19 post-compulsory course using ordinary least squares (OLS) regression with robust standard errors clustered at the secondary school level:

$$Y^*_{i,s,q} = \beta_0 + \sum\nolimits_{j=2}^{5} \beta_j (SES_i = j) + \delta' A'_i + \gamma' X'_i + \eta_s + \epsilon_{i,s,q}$$

where the variables are:

- $Y^*_{i,s,q}$: the measure of student-qualification match for individual i, who attended secondary school s and studied for post-compulsory qualification q
- $SES_i = j$: the quintile that each individual belongs to on the socio-economic index with the highest SES category used as the reference group
- A'_i: a vector of prior educational attainments of individual i at age 11, 14 and 16
- X'_i: a vector of several demographic characteristics for individual i
- η_s: the secondary school fixed effects
- $\epsilon_{i,s,q}$: the error component clustered at the secondary school level

I estimate the above equation examining the relationship between socio-economic quintile and academic match, sequentially by progressively increasing the number of covariates. In the first specification I include no additional covariates other than socio-economic status. In the second specification I introduce a set of prior attainment at ages 11, 14 and 16 for each student. In the third specification I introduce several demographic characteristics and in the final specification I add secondary school fixed effects. The coefficients in this final specification should be interpreted with caution. Although school characteristics might be important determinants of the post-compulsory education pathways the individuals choose to follow, the school a student attends is often determined by residential location and is therefore partly an outcome of socio-economic background[34]. As a consequence, the school fixed effects are likely to absorb some of the variation which should be attributed to socio-economic background and the resulting coefficients are downward biased.

[34] Gibbons and Machin (2003) indicated this relationship for primary schools in England by showing that a one percentage point increase in the neighbourhood proportion of children reaching the government-specified target grade pushes up neighbourhood property prices by 0.67%.

4.5 Empirical Results

In this section I indicate the relationship between socio-economic background (SES) and academic match in upper secondary education. In addition, I investigate the differences in the characteristics of the peers of the undermatched, matched and overmatched students by comparing means as an indication of how the school environments of undermatched, matched and overmatched students might differ. Exploring differences in the school environments the students are exposed to in more detail I estimate regressions with a full-set of prior attainment and demographic controls augmented by school composition variables. Further, I undertake a regional analysis showing the relationship between academic match and LAD's characteristics, including youth unemployment and the proportion of highly educated residents in each LAD. Finally, I examine the relationship between labour market income returns and academic assortative matching based on daily income acquired at age 25.

4.5.1 Main findings

The main findings identifying the relationship between socio-economic background (hereafter SES) and academic match are reported in Table 4.1. The first column shows the raw SES differences in academic match. Then, the extent to which these gaps can be accounted for by differences in other observable individual characteristics has been examined, including prior attainment at ages 11, 14 and 16 in column 2, basic demographics in column 3 and the secondary school the young person attended in column 4.

The findings reflect large differences in academic match by SES group. The first column showing the raw SES gradient in academic match is not very informative in providing insights about the direction of the relationship as the estimated coefficients take both negative and positive values and are even statistically insignificant for one of the SES groups. After accounting for prior attainment in column 2 the results start to become clear, showing a large and significant SES gradient in academic match. Since the matching index is negative when the students are undermatched, close to zero when they are matched and positive when they are overmatched, the negative and significant coefficients reported for students from all SES groups mean that they rank lower on the matching index than the students from the highest SES group[35]. For example, students who come from the lowest

[35]The negative coefficients do not necessarily indicate that the students are undermatched. They might still be overmatched to their upper-secondary qualification, just less than students from the Highest SES group.

SES group are found to average 0.16 standard deviations (s.d.) lower on the matching index compared to the students from the highest SES group. When introducing basic demographic characteristics in column 3 the SES gap observed in column 2 remains very similar. These findings suggest that differences in demographics from students of different SES groups provide only a minor explanation for why students from lower SES groups are less likely to rank higher on the matching index.

The final specification presented in column 4 adds secondary school fixed effects and the estimates, although decreasing substantially to about half, do not loose their precision. As mentioned above, the coefficients in this specification are likely to be downward biased but still, they do indicate that there is a significant relationship between SES and academic match which exists even between students of the same school.

To further understand these findings, it is important to examine how the identified SES gap varies across students of differing ability levels. In the section that follows the sample is divided into ability quintiles based on their performance in GCSE exams. The outcome of this analysis is discussed in Subsection 4.5.2. Further, the findings suggest that the secondary school attended absorbs a significant amount of the SES effect on academic match and therefore it would be very informative if we were able to disentangle between which features of the school are important in influencing academic match. Unfortunately, the available data do not allow such an investigation but, having information on several school characteristics, Subsection 4.5.3 attempts to provide an insight with regards to the large school gradient observed in the final specification.

The estimates on the effect of the other covariates are also of interest. The results suggest a significant gap between boys and girls with girls scoring about 0.04 s.d. higher on the matching index compared to boys in column 3 and about 0.03 s.d. higher in column 4, after applying school fixed effects. Also, the findings suggest a non-trivial impact of ethnic background with White students consistently found to score lower on the matching index compared to students from ethnic minority backgrounds.

In summary, the findings suggest that students from lower SES groups are more likely to score lower on the matching index compared to their more advantaged peers even when considering their prior attainment, demographics and secondary school attended. These findings are in line with the findings of other studies examining academic match in HE (Campbell et al., 2019, Hoxby and Avery, 2012, Smith et al., 2013) and provide evidence

that academic undermatch occurs before entrance to university.

4.5.2 Socio-economic effect on academic match across differing ability levels

Table 4.2 presents the results of the SES gap on academic match across ability quintiles based on the total GCSE point score of each individual[36]. Recall that in this analysis the individual GCSE point score is not included in the set of prior attainment controls.

Focusing on the most preferred specification (column 3) it can be observed that the SES gap becomes far more prominent across the highest achieving students. While the lowest achieving students from the lowest SES are found to average (surprisingly) about 0.01 s.d. higher on the matching index compared to the students from the highest SES, the highest achieving students from the lowest SES are found to average about 0.27 s.d. lower. The findings, therefore, suggest that among low ability students, SES makes little difference on students' match to upper-secondary qualifications while, on the other hand, among the high achieving students it is those from disadvantaged backgrounds who fail to match to their qualifications either by not seeing, understanding or wanting the opportunity that their ability opens for them.

The SES gap that is observed among high achieving students is non-trivial and these findings suggest that the SES effect works more by keeping high ability disadvantaged students down rather than keeping low ability advantaged students up. While this gap is substantially reduced when applying school fixed effects in column 4 the estimations still suggest that this gap exists even within schools. Interestingly, the SES gradient that is absorbed when school fixed effects are applied in the final specification increases dramatically among the highest achieving students. The findings suggest that among the highest achieving students almost 80% of the identified gap can be explained by differences in the schools that young individuals from differing backgrounds attend.

Comparing to the findings of Campbell et al. (2019) for students in HE, the SES gradient observed on academic match among the highest achieving students is of similar magnitude. For example, Campbell et al. (2019) reported that students from the lowest SES quintile average about 0.30 s.d. lower on the matching index compared to students from the highest

[36]As a robustness check the sample has been also divided by the number of A*-C GCSEs achieved (0 A*-C GCSEs, 1 A*-C GCSE, 2 A*-C GCSEs, 3 A*-C GCSEs, 4 A*-C GCSEs, 5 or more A*-C GCSEs without Maths and English, 5 or more A*-C GCSE including Maths and English) and the results remained qualitatively unchanged.

SES quintile, which is just about 0.03 s.d. greater than what is observed in upper-secondary education.

4.5.3 Analysis on school composition of undermatched, matched and overmatched students

The secondary school each student attended has been identified to explain a significant proportion of the SES gap in academic match. For this reason I take this investigation further and examine the differences in the schools that the undermatched, matched and overmatched students attended. The matching index suggests that students who are exactly average on the ability distribution should be matched to exactly average qualifications on the qualification-quality distribution. In this analysis, the students who matched have been considered those who scored between 0.5 s.d. below or above zero on the matching index. Therefore, the undermatched students are those who scored less than -0.5 s.d. on the matching index while the students who scored more than 0.5 s.d. on the matching index have been considered overmatched. In the sample, about 63% of the students are matched, 21% are undermatched and 16% are overmatched.

Table 4.3 presents the average school characteristics of the undermatched, matched and overmatched students by SES quintile. The available data do not allow us to disentangle the relative causality of these factors in producing academic mismatch but comparing means in this way provides an indication of how the environments of the students might be different. All the school characteristics are based on the secondary school the student attended at age 16.

The data suggest large differences in the school composition of students from different SES. The students from the lowest SES group are found in schools with higher proportions of peers who come from a disadvantaged background. For example students from the lowest SES group attend schools where the percentage of students with FSM eligibility is about 23%-25% while this proportion is decreased to just 7%-9% for students from the highest SES group. Further, students from higher SES groups attend schools with a greater proportion of high achieving peers, as identified from the percentage of students achieving 5 or more A*-C GCSEs including Maths and English, compared to students from lower SES groups. Also, students from lower SES groups are more likely to be found in schools with higher proportions of students coming from a non-White background.

Comparing within SES groups the observed differences between the undermatched, matched and overmatched students are less stark. However, it can be still observed that within all SES groups there tends to be a higher proportion of matched students in schools with higher proportions of high achieving students. Also, students who are both undermatched and overmatched to their qualification tend to be more apparent in schools with a greater proportion of disadvantaged students.

Table 4.4 explores the influence of these school characteristics in more detail by regressing the full set of SES indicators, prior attainment and demographic controls augmented with these school characteristics on the matching index. Differences between the environments of the undermatched, matched and overmatched students are still apparent. Students in schools with higher proportions of peers who are eligible for a FSM and who are from White-British ethnicities are less likely to score higher on the matching index while students in a school with higher proportions of high achieving peers are more likely to score higher. Table 4.5 shows the same information while splitting the sample into ability quintiles based on individual GCSE point score, instead of controlling for it. Interestingly, the negative effect of having more disadvantaged peers on academic match becomes greater in magnitude among the highest achieving students which is consistent with the main findings in Table 4.2. Similarly, the positive relationship between academic match and having a greater proportion of high achieving peers is stronger among the highest achieving students.

4.5.4 Does Local Authority District's characteristics impact academic match?

This subsection investigates the impact of LAD's characteristics on academic match regressing the regional characteristics, discussed in Subsection 4.3.5, on the academic matching index while controlling for SES, prior attainment and demographics[37]. The results from this analysis can be found in Table 4.6. The findings support what has been observed in the chloropleth maps. The students residing in urban areas are found to rank higher on the matching index compared to students residing in rural areas. Also, students who are residents in LADs which have a greater proportion of well educated residents are found to rank higher on the matching index. For example, a 10% increase in a LAD's

[37]I also run the same regressions while splitting the sample into ability quintiles based on individual GCSE point score, rather than controlling for it. The results follow the same pattern as the one observed for school composition characteristics presented in Table 4.5.

residents who have a Level 4 Qualification or above is associated with scoring about 0.03 s.d. higher on the matching index. Finally, students residing in areas with a greater proportion of youth unemployment are observed to rank lower on the matching index. Precisely, a 10% increase in the proportion of 16-24 claimants is associated with scoring about 0.06 s.d. lower on the matching index.

4.5.5 Income returns to academic assortative matching in upper-secondary education

This subsection estimates the returns to being academically assortatively matched using the method that has been suggested by Campbell et al. (2019) for HE and has been applied in this study for post-compulsory upper-secondary education. Making use of newly available data from HMRC tax records[38], I examine the effect of scoring higher on the matching index on the log daily income of students at age 25. Specifically, employer return forms (P14, P45 and P60) available in the dataset provide accurate information on earnings during the year (total annual pay in the 2015 tax year) and start and end date of periods of employment, for those who change employers during the year. These data were used to create a daily earnings measure, which is preferable to an annual earnings measure since it does not depend on the number of days worked per year, which will vary endogenously across individuals (McIntosh and Morris, 2018). Unfortunately no information on hours of work is included in the tax data, and so an hourly wage measure could not be obtained.

Further, the available data lack important information concerning when each individual entered the labour market. For example some of the students could enter the labour market after completing upper-secondary education when aged 18-19 and their income at age 25 includes 6-7 years of work experience. Other students studying for an undergraduate degree usually would enter the labour market at age 20-21 and those proceeding to a postgraduate degree would enter the labour market the earliest at age 21-22 and would have only 3-4 years of experience by age 25. I try to eliminate this bias by estimating separate models for university and non-university participants[39]. It should be acknowledged that those students who go to university are more likely to be higher achievers at age 16 and therefore more

[38]The matching was kindly undertaken by officials at the Department for Education, with the matched anonymised data set provided to us.

[39]A university participant is a student who has attended a course in any UK HE institution. The available data do not provide any additional information regarding the nature of the course or whether the individual has completed the course.

likely to be undermatched because of the mechanical relationship between high attainment at age 16 and chance of being undermatched. Therefore, in the analyses that follow it is vital to control for individual GCSE point score to avoid biased results with regards to income returns to academic match which should have been instead attributed to attainment at age 16. In any case, identifying a causal effect of being matched in upper-secondary education on future returns is beyond the scope of this study. The analysis attempts to indicate the direction of this relationship rather than to assign a causality to it.

The outcome of this analysis is presented in Table 4.7. Due to the well-established differences in returns to education by gender in addition to the aggregated sample results the results are also presented separately for boys and girls. The results indicate a positive relationship between log daily income at age 25 and scoring higher on the matching index, especially for non-university participants. For example, the aggregated results for students who did not attend a HE course indicate that scoring 1 s.d. higher on the matching index is associated with higher log daily income of 0.09 points, which is approximately equal to a 9% increase in actual income. Comparing between boys and girls, the returns to academic assortative matching are much more apparent for girls. For university participants, scoring a 1 s.d. higher on the matching index is associated with about a 9% increase in log daily income at age 25 for girls while for boys it is associated with a 2% decrease on log daily income. For non-university participants the effect is still notably higher for girls. Girls who matched 1 s.d. higher on the matching index are found to earn about 18% more compared to boys who earned about 2% more.

The results of this analysis for the other variables are also of interest and are therefore presented in Appendix C.3. The main findings of this paper suggested that girls and students from ethnic minorities were the ones identified to score higher on the matching index. The current analysis established the positive relationship between income and academic assortative matching. It is interesting and disappointing, hence, to see that there is a significant gender and ethnic gap on income returns with girls and individuals from ethnic minority backgrounds earning significantly less, although being the ones who scored higher on the matching index (though that is held constant in this analysis). Especially for non-university participants the gender gap is non-trivial. Girls are found to earn 36% less than boys. Also non-university participating Black students are earning 24% less than White students.

Evidently, these findings indicate a significant advantage of scoring higher on the matching index on income returns, at least at early ages. Analysing income returns using the matching index assumes that its effect is linear while it could be that it is being matched (scoring in the middle of the index) that has positive income returns rather than being both overmatched or undermatched. On the other hand, it might be the case that it is the overmatched students that are driving these positive returns on income. For that reason, one additional analysis has been carried out to enlighten with regards to this matter. The matching index has been recoded to take a value of 0 for all the individuals who scored at or above -0.5 (all matched and overmatched students). This variable will show the income returns of being 1 s.d. less undermatched. Then a second variable has been created for the overmatched students, recoding the matching index to take a value of 0 for all students who scored at or below 0.5 (all the matched and undermatched students). The coefficients of this variable will indicate the income returns of being 1 s.d. more overmatched. As before separate results are presented for university and non-university participants and for girls and boys.

The outcome of this analysis is presented in Table 4.8. The effect of being less undermatched on labour market income returns is now more apparent and still considerably greater in magnitude for girls than boys. Non-university participating girls who were 1 s.d. less undermatched are found to earn about 17% more. For university participating girls this effect decreases to 10% but it still remains positive and significant. The effect of academic undermatch has also become more apparent for boys although it is smaller in magnitude compared to girls. University participating boys who were 1 s.d. less undermatched in their upper-secondary qualification earned 3% more at age 25 while non-university participants earned about 5% more.

Turning to the effect of being more overmatched, in order to clarify whether the earlier findings are driven by students who are overmatched, we can see that this is not the case. For university participating students, more overmatching is associated with a negative daily income return. For non-university participants, there is a positive return to being more overmatched but it is much smaller than the effect of being less undermatched. One possible explanation that we could assign to this outcome is that students pursuing qualifications which are beyond their academic credentials could be achieving them with a very low grade.

The outcome of this analysis highlights the importance of being academically assortatively matched, establishing the strong and positive relationship of being matched and labour market income returns. The results indicate that this positive effect is not driven by students overmatching to qualifications but that it is actually an outcome of being academically matched. This is evident from the fact that while being less undermatched has a significant positive impact on labour market returns, being more overmatched does not have an equally positive effect.

In order to convince on the robustness of these findings there is one potential parameter that needs to be considered. Given how the matching index is constructed, $Match_{i,q} = med(zgcse_q) - zgcse_i$, when controlling for individual GCSE point score in the estimated models it is possible that the positive coefficients observed in Table 4.7, instead of reflecting positive income returns to scoring higher on the matching index rather reflect positive income returns to achieving a higher upper-secondary qualification which would thus result in a higher score on the matching index, given that the individual GCSE point score is kept constant. I, therefore, carry out one additional analysis to examine the robustness of the results.

A categorical variable has been constructed identifying whether the individual is matched (scoring between -0.5 s.d. and 0.5 s.d. on the matching index), undermatched (scoring below -0.5 s.d. on the matching index) or overmatched (scoring above 0.5 s.d. on the matching index) and the continuous measure of academic match has been replaced with this new categorical variable. This robustness check is important as it allows us to control for both the individual GCSE point score and the upper-secondary qualification achieved by the individual and therefore any observed differences in the estimated coefficients will reflect the effect of being matched rather than being undermatched and being matched rather than being overmatched while keeping these two important parameters constant. The outcome of this analysis is presented in Appendix C.4. The results show a negative log income return to being both overmatched and undermatched compared to matched individuals suggesting the robustness of the main estimations.

4.6 Conclusion

The chapter uncovers a significant socio-economic gap in academic match among English students in upper-secondary post-compulsory education. Students from lower socio-economic

backgrounds achieve less highly ranked qualifications compared to their similarly attaining but more advantaged peers. This gap is identified to be greater among the highest achieving students suggesting that socio-economic inequality works more by keeping high ability disadvantaged students down than keeping low ability advantaged students up. The analysis reveals that students from low socio-economic backgrounds attend disadvantaged schools with significantly lower proportions of high achieving peers and greater proportions of disadvantaged peers and that among the highest achieving students 80% of the identified socio-economic gap can be explained by such differences in the schools that young people from differing backgrounds attend. Further, the study reveals that geographical factors are also strong predictors of academic match with the most undermatched students being found in mostly rural areas with greater proportions of youth unemployment and lower proportions of highly educated residents. It is also demonstrated that there is a significant labour market cost of being undermatched, especially for non-university participating girls who are found to earn 17% more at age 25 when being 1 s.d. less undermatched.

Hupkau et al. (2017) identified that students undertaking vocational qualifications at Level 2 or below are much more likely to be from disadvantaged family backgrounds than those undertaking higher levels of qualifications and that generally they do not have the pre-requisites to start their post-compulsory education at a higher level. This study reveals that disadvantaged students, even if qualified to study for a higher level qualification, are still more likely to undermatch in less highly ranked qualifications. There are important policy implications to be drawn from these findings. Policy-makers interested in social mobility should be focusing more on providing students with information related to the available upper-secondary courses that are suitable to each student's ability credentials and future educational and occupational aspirations.

In the future I plan to expand this analysis by examining whether increased provision of information, advice and guidance in schools about available post-compulsory education routes is efficient in decreasing academic mismatch in upper-secondary education. Further, the role of educational and occupational aspirations in influencing academic match would be an interesting area to investigate. Unfortunately, the administrative data used in this study do not provide such information. For future research rich individual level survey data will be used to expand this investigation and study whether the lack of educational advice or the heterogeneous educational and professional aspirations among students are driving some of the socio-economic effect that has been identified in this analysis.

Table 4.1: Socio-economic effect on academic match in post-compulsory education

	Specification (1)	Specification (2)	Specification (3)	Specification (4)
Socio-economic background				
Middle-High SES	-0.015***	-0.052***	-0.049***	-0.026***
	(0.006)	(0.003)	(0.002)	(0.002)
Middle SES	0.000	-0.086***	-0.082***	-0.048***
	(0.007)	(0.003)	(0.003)	(0.002)
Middle-Low SES	0.030***	-0.123***	-0.119***	-0.071***
	(0.008)	(0.004)	(0.003)	(0.002)
Lowest SES	0.083***	-0.159***	-0.161***	-0.087***
	(0.010)	(0.005)	(0.004)	(0.003)
Control Variables				
KS2 test scores		-0.030***	-0.033***	-0.036***
		(0.002)	(0.002)	(0.001)
KS3 level scores		0.295***	0.232***	0.193***
		(0.003)	(0.003)	(0.002)
GCSE point score		-0.717***	-0.775***	-0.718***
		(0.004)	(0.003)	(0.003)
5≥ A*-C GCSEs including Maths & English			0.287***	0.258***
			(0.003)	(0.003)
Girl			0.036***	0.027***
			(0.002)	(0.002)
Asian			0.228***	0.193***
			(0.005)	(0.004)
Black			0.164***	0.117***
			(0.006)	(0.005)
Other			0.194***	0.144***
			(0.010)	(0.008)
Mixed			0.060***	0.039***
			(0.005)	(0.004)
Observations	391,651	391,651	391,651	391,651
Number of clusters	3,076	3,076	3,076	3,076
R-squared	0.00	0.58	0.61	0.64
School FEs				✓

Notes: Post-compulsory education participants only. Omitted groups: Highest SES, Boy, White.
Standard errors clustered at secondary school level and reported in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, **** $p < 0.01$.

Table 4.2: Socio-economic effect on academic match in post-compulsory education across ability quintiles

	Specification (1)	Specification (2)	Specification (3)	Specification (4)
Bottom Quintile of GCSE scores				
Middle-High SES	-0.032***	-0.038***	-0.036***	-0.025***
	(0.008)	(0.008)	(0.008)	(0.008)
Middle SES	-0.020***	-0.032***	-0.030***	-0.019**
	(0.007)	(0.008)	(0.007)	(0.008)
Middle-Low SES	0.000	-0.023***	-0.022***	-0.014*
	(0.008)	(0.008)	(0.008)	(0.008)
Lowest SES	0.055***	0.014*	0.015*	0.023***
	(0.008)	(0.008)	(0.008)	(0.008)
Observations	97,909	97,909	97,909	97,909
Number of clusters	2,989	2,989	2,989	2,989
R-squared	0.00	0.03	0.03	0.11
Middle-Low Quintile of GCSE scores				
Middle-High SES	-0.059***	-0.050***	-0.049***	-0.024***
	(0.006)	(0.005)	(0.005)	(0.004)
Middle SES	-0.091***	-0.075***	-0.077***	-0.041***
	(0.006)	(0.006)	(0.005)	(0.005)
Middle-Low SES	-0.126***	-0.098***	-0.106***	-0.057***
	(0.007)	(0.006)	(0.006)	(0.005)
Lowest SES	-0.200***	-0.146***	-0.168***	-0.082***
	(0.009)	(0.008)	(0.007)	(0.006)
Observations	99,120	99,120	99,120	99,120
Number of clusters	3,047	3,047	3,047	3,047
R-squared	0.02	0.07	0.12	0.23
Middle-High Quintile of GCSE scores				
Middle-High SES	-0.064***	-0.042***	-0.040***	-0.014***
	(0.005)	(0.004)	(0.004)	(0.003)
Middle SES	-0.117***	-0.078***	-0.074***	-0.033***
	(0.006)	(0.005)	(0.004)	(0.004)
Middle-Low SES	-0.187***	-0.123***	-0.118***	-0.057***
	(0.007)	(0.006)	(0.005)	(0.004)
Lowest SES	-0.294***	-0.178***	-0.173***	-0.076***
	(0.010)	(0.008)	(0.006)	(0.006)
Observations	96,398	96,398	96,398	96,398
Number of clusters	3,051	3,051	3,051	3,051
R-squared	0.04	0.18	0.24	0.35
Highest Quintile of GCSE scores				
Middle-High SES	-0.078***	-0.053***	-0.056***	-0.005
	(0.008)	(0.007)	(0.007)	(0.004)
Middle SES	-0.142***	-0.096***	-0.099***	-0.018***
	(0.011)	(0.010)	(0.010)	(0.005)
Middle-Low SES	-0.227***	-0.151***	-0.155***	-0.020***
	(0.013)	(0.011)	(0.011)	(0.005)
Lowest SES	-0.398***	-0.266***	-0.266***	-0.040***
	(0.019)	(0.015)	(0.014)	(0.007)
Observations	98,224	98,224	98,224	98,224
Number of clusters	3,010	3,010	3,010	3,010
R-squared	0.05	0.10	0.12	0.43
Prior attainment (KS2, KS3)		✓	✓	✓
Demographics			✓	✓
School FEs				✓

Notes: Post-compulsory education participants only.
Omitted group: Highest SES. Standard errors clustered at secondary school level and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.3: School composition of the undermatched, matched and overmatched students by socio-economic group

	Undermatched	Matched	Overmatched
Highest SES			
% with FSM eligibility	7.76	7.03	9.22
	(7.78)	(7.14)	(9.44)
% achieving 5≥ A*-C GCSEs including Maths & English	58.13	58.59	52.81
	(18.79)	(17.73)	(16.59)
% of White British	83.41	83.33	79.73
	(18.49)	(17.99)	(22.00)
Observations	15,974	60,669	9,855
Middle-High SES			
% with FSM eligibility	9.40	8.54	10.63
	(8.11)	(7.48)	(9.14)
% achieving 5≥ A*-C GCSEs including Maths & English	52.26	53.19	47.08
	(19.16)	(18.38)	(16.26)
% of White British	85.30	84.68	81.75
	(18.66)	(18.60)	(21.73)
Observations	16,981	56,458	10,160
Middle SES			
% with FSM eligibility	12.01	10.88	13.21
	(9.56)	(9.10)	(10.52)
% achieving 5≥ A*-C GCSEs including Maths & English	46.62	48.04	42.08
	(18.66)	(18.53)	(15.85)
% of White British	85.09	83.88	80.63
	(19.63)	(20.84)	(24.01)
Observations	16,797	51,796	11,534
Middle-Low SES			
% with FSM eligibility	16.06	14.38	17.14
	(11.50)	(11.19)	(12.09)
% achieving 5≥ A*-C GCSEs including Maths & English	40.90	42.93	37.12
	(17.73)	(17.80)	(15.02)
% of White British	83.18	81.74	78.18
	(22.45)	(23.81)	(26.32)
Observations	15,956	44,735	13,376
Lowest SES			
% with FSM eligibility	24.78	22.23	23.86
	(14.49)	(14.07)	(13.69)
% achieving 5≥ A*-C GCSEs including Maths & English	33.05	35.23	30.95
	(15.79)	(15.96)	(13.51)
% of White British	80.13	77.19	77.32
	(25.68)	(28.08)	(27.25)
Observations	15,118	35,906	16,336

Note: The numbers presented in each column are the mean values of each school characteristic for post-compulsory education participants in the analytical sample. The numbers in parentheses are the standard deviations.

Table 4.4: School composition and academic match

	(1)	(2)	(3)
% with FSM eligibility	-0.002***		
	(0.000)		
% achieving 5≥ A*-C GCSEs including Maths & English		0.002***	
		(0.000)	
% of White British			-0.001***
			(0.000)
Observations	391,651	391,651	391,651
Number of clusters	3,076	3,076	3,076
R-squared	0.61	0.62	0.61
SES	✓	✓	✓
Prior achievement (KS2, KS3, KS4)	✓	✓	✓
Demographics	✓	✓	✓

Notes: Post-compulsory education participants only.
Standard errors clustered at secondary school level and reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.5: School composition and academic match across ability quintiles

	(1) Bottom Quintile	(2) Middle-Low Quintile	(3) Middle-High Quintile	(4) Highest Quintile
% with FSM eligibility	0.001***	0.002***	-0.002***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.001)
Observations	97,909	99,120	96,398	98,224
Number of clusters	2,989	3,047	3,051	3,010
R-squared	0.03	0.07	0.25	0.13
% achieving 5≥ A*-C GCSEs	-0.001***	0.001***	0.002***	0.003***
including Maths & English	(0.000)	(0.000)	(0.000)	(0.000)
Observations	97,909	99,120	96,398	98,224
Number of clusters	2,989	3,047	3,051	3,010
R-squared	0.03	0.07	0.25	0.13
% of White British	-0.001***	-0.003***	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	97,909	99,120	96,398	98,224
Number of clusters	2,989	3,047	3,051	3,010
R-squared	0.03	0.09	0.25	0.12
SES	✓	✓	✓	✓
Prior achievement (KS2, KS3)	✓	✓	✓	✓
Demographics	✓	✓	✓	✓

Notes: Post-compulsory education participants only. Ability quintiles have been calculated based on individual GCSE point scores. Standard errors clustered at secondary school level and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.6: Local Authority District (LAD) characteristics and academic match

	(1)	(2)	(3)
Urban LAD	0.001		
	(0.004)		
% with L4 Qualifications or above		0.003***	
		(0.000)	
% of 16-24 claimants			-0.006***
			(0.001)
Observations	391,651	391,651	391,651
Number of clusters	3,076	3,076	3,076
R-squared	0.62	0.62	0.62
SES	✓	✓	✓
Prior achievement (KS2, KS3, KS4)	✓	✓	✓
Demographics	✓	✓	✓

Notes: Post-compulsory education participants only.
Standard errors clustered at secondary school level and reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.7: Log income returns to academic assortative matching

	Full sample		Girls		Boys	
	University	No University	University	No University	University	No University
	(1)	(2)	(3)	(4)	(5)	(6)
Academic match	0.035***	0.088***	0.088***	0.182***	-0.022**	0.019***
	(0.007)	(0.005)	(0.010)	(0.008)	(0.011)	(0.007)
Observations	164,259	143,901	90,951	65,626	73,308	78,275
Number of clusters	3,061	3,052	2,879	2,874	2,822	2,831
R-squared	0.02	0.09	0.02	0.07	0.02	0.05
SES	✓	✓	✓	✓	✓	✓
Prior achievement (KS2, KS3, KS4)	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓

Notes: Post-compulsory education participants matched to HMRC tax records only.
Standard errors clustered at secondary school level and reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.8: Log income returns to being less undermatched and more overmatched

	Full sample		Girls		Boys	
	University	No University	University	No University	University	No University
	(1)	(2)	(3)	(4)	(5)	(6)
Less undermatched	0.064***	0.101***	0.095***	0.170***	0.026**	0.048***
	(0.008)	(0.006)	(0.010)	(0.009)	(0.013)	(0.008)
More overmatched	-0.052***	0.040***	-0.014	0.112***	-0.086***	-0.008
	(0.010)	(0.008)	(0.014)	(0.012)	(0.015)	(0.010)
Observations	164259	143901	90951	65626	73308	78275
Number of clusters	3,061	3,052	2,879	2,874	2,822	2,831
R-squared	0.02	0.09	0.02	0.07	0.02	0.05
SES	✓	✓	✓	✓	✓	✓
Prior achievement (KS2, KS3, KS4)	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓

Notes: Post-compulsory education participants matched to HMRC tax records only.
Standard errors clustered at secondary school level and reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C:

C.1 Comparison of total and analytical sample

Variable	Mean/ (sd)	N	Mean/ (sd)	N	Difference
Highest SES	0.20 (0.40)	555,601	0.22 (0.41)	391,651	-0.02(0.00)***
Middle-High SES	0.20 (0.40)	555,601	0.21 (0.41)	391,651	-0.1(0.00)***
Middle SES	0.20 (0.40)	555,601	0.20 (0.40)	391,651	-0.00(0.00)***
Middle-Low SES	0.20 (0.40)	555,601	0.19 (0.39)	391,651	0.01(0.00)***
Lowest SES	0.20 (0.40)	555,601	0.17 (0.38)	391,651	0.03(0.00)***
KS2 scores (standardised)	0.00 (1.00)	525,229	0.13 (0.97)	391,651	-0.13(0.00)***
KS3 level scores (standardised)	0.00 (1.00)	537,332	0.18 (0.95)	391,651	-0.18(0.00)***
GCSE point score (standardised)	-0.00 (1.00)	555,601	0.22 (0.90)	391,651	-0.22(0.00)***
5≥ A*-C GCSEs including Maths & English	0.45 (0.50)	555,601	0.54 (0.50)	391,651	-0.09(0.00)***
Female	0.50 (0.50)	555,601	0.51 (0.50)	391,651	-0.01(0.00)***
White	0.86 (0.34)	541,210	0.88 (0.33)	391,651	0.00(0.00)***
Asian	0.07 (0.25)	541,210	0.06 (0.25)	391,651	0.00(0.00)***
Black	0.04 (0.19)	541,210	0.03 (0.17)	391,651	0.00(0.00)***
Other	0.01 (0.09)	541,210	0.01 (0.08)	391,651	0.00(0.00)***
Mixed	0.02 (0.15)	541,210	0.02 (0.15)	391,651	0.00(0.00)**
Q1: L1 Vocational	0.09 (0.29)	428,802	0.08 (0.27)	391,651	0.01(0.00)***
Q2: L2 Vocational	0.21 (0.40)	428,802	0.20 (0.40)	391,651	0.01(0.00)**
Q3: L3 Vocational	0.18 (0.38)	428,802	0.18 (0.39)	391,651	-0.00(0.00)***
Q4: AS levels	0.07 (0.26)	428,802	0.08 (0.26)	391,651	-0.01(0.00)
Q5: A levels<3 w/o facilitating subjects	0.09 (0.28)	428,802	0.09 (0.28)	391,651	-0.00(0.00)***
Q6: A levels<3 with 1 facilitating subject	0.04 (0.19)	428,802	0.04 (0.19)	391,651	0.00(0.00)
Q7: A levels≥3 w/o facilitating subjects	0.07 (0.26)	428,802	0.07 (0.26)	391,651	-0.00(0.00)***
Q8: 2 A levels in 2 facilitating subjects	0.01 (0.09)	428,802	0.01 (0.09)	391,651	0.00(0.00)
Q9: A levels≥3 with 1 facilitating subject	0.10 (0.30)	428,802	0.10 (0.30)	391,651	-0.00(0.00)***
Q10: A levels≥3 with 2 facilitating subjects	0.08 (0.27)	428,802	0.08 (0.28)	391,651	-0.00(0.00)*
Q11 :A levels≥3 with ≥3	0.06 (0.24)	428,802	0.06 (0.24)	391,651	0.00(0.00)

The last column presents the t-test estimates for mean differences of each variable in the total and analytical sample. Standard errors are presented in parentheses. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.2 Descriptive statistics

SES Quintile	(1) Highest SES	(2) Middle-High SES	(3) Middle SES	(4) Middle-Low SES	(5) Lowest SES	Total
Prior achievement (standardised)						
KS2 scores	0.44	0.28	0.15	-0.05	-0.29	0.13
	(0.87)	(0.91)	(0.94)	(0.99)	(1.02)	(0.97)
KS3 level scores	0.54	0.37	0.20	-0.02	-0.35	0.18
	(0.85)	(0.88)	(0.90)	(0.93)	(0.95)	(0.95)
GCSE point score	0.50	0.38	0.25	0.07	-0.18	0.22
	(0.78)	(0.82)	(0.85)	(0.91)	(1.00)	(0.90)
Demographics						
5≥ A*-C GCSEs including Maths & English	0.71	0.63	0.55	0.44	0.31	0.54
	(0.45)	(0.48)	(0.50)	(0.50)	(0.46)	(0.50)
Girl	0.50	0.51	0.51	0.51	0.51	0.51
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
White	0.91	0.91	0.89	0.86	0.81	0.88
	(0.28)	(0.29)	(0.31)	(0.35)	(0.40)	(0.33)
Asian	0.04	0.05	0.06	0.07	0.11	0.06
	(0.20)	(0.21)	(0.23)	(0.26)	(0.32)	(0.25)
Black	0.02	0.02	0.03	0.04	0.04	0.03
	(0.13)	(0.14)	(0.16)	(0.20)	(0.21)	(0.17)
Other	0.01	0.01	0.01	0.01	0.01	0.01
	(0.08)	(0.07)	(0.07)	(0.09)	(0.08)	(0.08)
Mixed	0.02	0.02	0.02	0.02	0.03	0.02
	(0.15)	(0.14)	(0.14)	(0.15)	(0.16)	(0.15)
Upper-secondary qualification (ordered)						
Q1: L1 Vocational	0.03	0.05	0.07	0.11	0.17	0.08

Table C.1 –continued on next page

Table C.1 – continued from previous page

	(0.18)	(0.22)	(0.25)	(0.31)	(0.38)	(0.27)
Q2: L2 Vocational	0.11	0.16	0.20	0.25	0.32	0.20
	(0.32)	(0.37)	(0.40)	(0.44)	(0.47)	(0.40)
Q3: L3 Vocational	0.14	0.18	0.20	0.20	0.20	0.18
	(0.35)	(0.38)	(0.40)	(0.40)	(0.40)	(0.39)
Q4: AS levels	0.07	0.08	0.08	0.08	0.07	0.08
	(0.26)	(0.27)	(0.27)	(0.27)	(0.25)	(0.26)
Q5: A levels<3 w/o facilitating subjects	0.09	0.09	0.09	0.09	0.07	0.09
	(0.29)	(0.29)	(0.29)	(0.28)	(0.26)	(0.28)
Q6: A levels<3 with 1 facilitating subject	0.04	0.04	0.04	0.04	0.03	0.04
	(0.20)	(0.20)	(0.20)	(0.19)	(0.16)	(0.19)
Q7: A levels≥3 w/o facilitating subjects	0.10	0.09	0.07	0.06	0.04	0.07
	(0.30)	(0.28)	(0.26)	(0.24)	(0.20)	(0.26)
Q8: 2 A levels in 2 facilitating subjects	0.01	0.01	0.01	0.01	0.01	0.01
	(0.09)	(0.09)	(0.09)	(0.09)	(0.07)	(0.09)
Q9: A levels≥3 with 1 facilitating subject	0.15	0.12	0.10	0.07	0.05	0.10
	(0.36)	(0.33)	(0.30)	(0.26)	(0.21)	(0.30)
Q10: A levels≥3 with 2 facilitating subjects	0.13	0.10	0.08	0.06	0.03	0.08
	(0.34)	(0.30)	(0.27)	(0.23)	(0.17)	(0.28)
Q11 :A levels≥3 with ≥3 facilitating subjects	0.11	0.08	0.06	0.04	0.02	0.06
	(0.31)	(0.26)	(0.23)	(0.19)	(0.14)	(0.24)
Observations	86,498	83,599	80,127	74,067	67,360	391,651

Note: The numbers presented in each column are the mean values of each characteristic for post-compulsory education participants in the analytical sample. The numbers in parentheses are the standard deviations.

C.3 Log Income returns to academic assortative matching

	All sample		Girls		Boys	
	University	No University	University	No University	University	No University
	(1)	(2)	(3)	(4)	(5)	(6)
Academic match	0.035***	0.088***	0.088***	0.182***	-0.022**	0.019***
	(0.007)	(0.005)	(0.010)	(0.008)	(0.011)	(0.007)
Middle-High SES	-0.010	-0.006	-0.008	-0.012	-0.013	-0.001
	(0.006)	(0.007)	(0.008)	(0.010)	(0.009)	(0.009)
Middle SES	-0.037***	-0.018**	-0.038***	-0.030***	-0.035***	-0.008
	(0.007)	(0.007)	(0.009)	(0.010)	(0.010)	(0.009)
Middle-Low SES	-0.053***	-0.061***	-0.055***	-0.093***	-0.050***	-0.034***
	(0.007)	(0.007)	(0.009)	(0.011)	(0.011)	(0.009)
Lowest SES	-0.113***	-0.133***	-0.110***	-0.160***	-0.117***	-0.109***
	(0.008)	(0.007)	(0.011)	(0.011)	(0.013)	(0.010)
KS2 scores	0.030***	0.031***	0.033***	0.021***	0.027***	0.039***
	(0.005)	(0.004)	(0.007)	(0.006)	(0.008)	(0.006)
KS3 Level scores	0.009	0.034***	-0.010	0.047***	0.031***	0.024***
	(0.007)	(0.005)	(0.008)	(0.008)	(0.010)	(0.007)
GCSE point score	0.089***	0.190***	0.132***	0.266***	0.045***	0.137***
	(0.008)	(0.006)	(0.010)	(0.009)	(0.012)	(0.007)
5≥ A*-C GCSEs including Maths & English	0.125***	0.040***	0.138***	0.048***	0.111***	0.026***
	(0.007)	(0.006)	(0.009)	(0.009)	(0.010)	(0.008)
Female	-0.048***	-0.355***				
	(0.005)	(0.005)				
Asian	0.056***	-0.116***	0.043***	0.039*	0.070***	-0.235***
	(0.009)	(0.014)	(0.011)	(0.021)	(0.012)	(0.019)
Black	-0.078***	-0.236***	-0.078***	-0.082***	-0.082***	-0.348***
	(0.013)	(0.019)	(0.016)	(0.026)	(0.021)	(0.025)

Table C.3.1 –continued on next page

Table C.3.1 – continued from previous page

Other	-0.009	-0.142***	0.003	0.002	-0.023	-0.264***
	(0.026)	(0.045)	(0.031)	(0.056)	(0.040)	(0.064)
Mixed	-0.062***	-0.094***	-0.062***	-0.042*	-0.063***	-0.141***
	(0.015)	(0.017)	(0.020)	(0.025)	(0.023)	(0.024)
Observations	164,259	143,901	90,951	65,626	73,308	78275
Number of clusters	3,061	3,052	2,879	2,874	2,822	2,831
R-squared	0.02	0.09	0.02	0.07	0.02	0.05

Notes: Post-compulsory education participants matched to HMRC tax records only. Omitted groups: Highest SES, Boy, White.

Standard errors clustered at secondary school level and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.4 Log income returns to academic assortative matching (categorical variable)

	Full sample		Girls		Boys	
	University	No University	University	No University	University	No University
	(1)	(2)	(3)	(4)	(5)	(6)
Undermatched	-0.013	-0.002	-0.017	0.007	-0.005	-0.009
	(0.009)	(0.008)	(0.011)	(0.012)	(0.014)	(0.011)
Overmatched	-0.067***	-0.041***	-0.056***	-0.033***	-0.073***	-0.044***
	(0.009)	(0.008)	(0.012)	(0.012)	(0.013)	(0.011)
Observations	164,259	143,901	90,951	65,626	73,308	78,275
Number of clusters	3,061	3,052	2,879	2,874	2,822	2,831
R-squared	0.02	0.10	0.02	0.07	0.02	0.06
SES	✓	✓	✓	✓	✓	✓
Upper-secondary qualification	✓	✓	✓	✓	✓	✓
Prior achievement (KS2, KS3, KS4)	✓	✓	✓	✓	✓	✓
Demographics	✓	✓	✓	✓	✓	✓

Notes: Post-compulsory education participants matched to HMRC tax records only.
Standard errors clustered at secondary school level and reported in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 5

Conclusions

5.1 Motivations and aims

Each empirical chapter of this thesis presented robust empirical evidence from large-scale data sources with the aim to inform on the posed questions, in three related aspects of the Economics of Education, concerning the determinants of the educational aspirations and choices of young people. Throughout this thesis, educational aspirations have been defined as individuals' future plans while choices were the actual outcomes that have been observed, based on achieved qualifications, and have been defined as the selection between academic, vocational and no post-compulsory education. Evaluating the schooling system with regards to the factors influencing the post-compulsory educational decisions of young people is crucial, particularly when considering the worldwide expansion of post-compulsory education.

It would be safe to say that individual ability, cognitive skills and IQ are what determines the educational outcomes of students and that holding these constant, then a meritocratic educational system, such as the one England, should provide equality of opportunity (Galindo Rueda and Vignoles, 2003). However, this is not the case for two main reasons. The first, being examined by related literature, concerns the timing of the emergence of differences in the cognitive development of different groups of children. For England, Feinstein (2003) showed that differences in educational achievement between socio-economic groups emerge early in pre-school and primary school, suggesting that socio-economic background is directly affecting the cognitive development of students from very early ages. The study uses UK data to show that there is a large decline in test performance of disadvantaged students between 22 months old and 10 years old. The findings of this study, and of several subsequent studies which used the same methodology, including Schoon (2006), Blanden and Machin (2007, 2010) and Parsons et al. (2011), have been criticised by Jerrim and Vignoles (2013) as not taking into account the statistical artefact known as "regression to the mean" and therefore producing biased results. Nevertheless, although Jerrim and Vignoles (2013) when addressing this problem with an alternative methodology

no longer find much significant evidence that able but disadvantaged students fall behind their more advantaged but less able peers, they still confirm that socio-economic gaps in children's test scores are large and apparent from a very early age.

Secondly, educational outcomes of young people are proven to be influenced by a myriad of other factors other than individual intelligence. This thesis has established that, even when conditioning on exceptionally detailed measures of individual ability, two other very important aspects are influencing educational aspirations and choices; namely socio-economic background -directly and not through its impact on cognitive development- and school peers' characteristics. Further, although beyond the principal aim of this thesis, the empirical analyses provide evidence of other important drivers of educational outcomes, those being the composition of the school the student attends and demographic characteristics including gender and ethnic background.

Considering the aims of this thesis in more detail, the first empirical study, in Chapter 2, investigated whether the importance of socio-economic background and ability in determining the post-compulsory educational aspirations and choices of young people has changed over time. The chapter used unique longitudinal data sets of two English cohorts, the older born in a single week of 1970 using data from the British Cohort Study (BCS), and the more recent born between 1989-1990 using data from the Longitudinal Study of Young People in England (LSYPE). Adopting multinomial logistic techniques and a within-and-between cohorts comparison framework, the chapter evaluated the changing influence of socio-economic background and ability from the older to the more recent cohort as well as the differences of these effects between aspirations and choices within the sample of the same cohort.

The second empirical study, presented in Chapter 3, investigated whether the educational aspirations of the LSYPE cohort, analysed in the previous chapter, are influenced by the characteristics of their secondary school peers. These characteristics include peers' ability, peers' socio-economic background and peers' own aspirations. In order to overcome the endogeneity and selection biases associated with peer effects, the study adopted an identification strategy based on 'peers-of-peers'. Specifically, each individual's secondary school peers were instrumented with their primary school peers who did not attend the same primary or secondary school as the individual. These peers-of-peers will have affected the secondary school peers through attendance at the same primary school, but have likely

never met the individual and therefore did not have any direct effect on the individual's aspirations.

The final empirical study, presented in Chapter 4, used detailed administrative records from schools and tax authorities in England for the whole population of the same cohort, born between 1989-1990, to investigate the impact of students' socio-economic background on their academic match in 16-19 post-compulsory education. Academic match was determined using a continuous measure of student-qualification match which identified undermatched, matched and overmatched students. The selectivity of each upper secondary qualification was classified from the median age 16 exam score achievements of the students studying for that qualification. Then, each student's ability position on the age 16 test score distribution was compared, examining whether the students matched to qualifications with similarly-attaining peers.

5.2 Summary of results

The results of the first empirical study, presented in Chapter 2, suggest that there are decreasing socio-economic and ability effects on both aspirations and choices for academic and no post-compulsory education and provide evidence that the expansion of academic education has proportionately benefited individuals from all social backgrounds. Further, the study identified a decline in vocational education participation which did not arise from falling aspirations but because of rising aspirations and actual participation in academic education.

The results of the second study, presented in Chapter 3, suggest that peer effects on individuals' intentions to stay in education are insignificant for girls but not for boys. Conditional on students' plans to remain in post-compulsory education, peers' ability, socio-economic background and aspirations to follow an academic rather than a vocational education pathway, are all identified to have a positive and significant effect on individuals' aspirations to follow an academic route. The study also showed evidence that the provision of information, advice and guidance by schools or external agencies can serve to mitigate peer effects. Finally, individuals with higher ability and more socially-advantaged peers were identified to be less likely to have changed their educational aspirations between Year 9 and Year 11 of schooling.

The results of the final study, discussed in Chapter 4, suggest that disadvantaged students

are more likely to be exposed to academic undermatch compared to their more advantaged peers and that the phenomenon is still apparent even within students of the same school. The study also identified that undermatched students are more likely to be found in schools with lower proportions of high achieving students and higher proportions of ethnic minority and disadvantaged students. In addition, the study identified indications that critical masses of undermatched students are more likely to be found in rural districts with higher rates of youth unemployment and higher proportions of poorly educated residents. Finally, the study demonstrated that academic assortative matching has a positive relationship with labour market returns, at least at early ages.

Overall, the empirical results of this thesis support that young people's educational decisions are socially graded, with students from disadvantaged backgrounds consistently found to aspire and to choose different post-compulsory educational routes compared to their more advantaged peers. This thesis showed that this social bias has been significantly reduced over time, yet has not disappeared. The post-compulsory qualifications achieved by disadvantaged youth are found to be undermatched to their academic credentials causing severe costs on their labour market income returns. Further, this thesis has established the significance of secondary school peers in influencing the post-compulsory educational aspirations of young people and that the provision of information, advice and guidance can serve to mitigate the importance of peers in influencing these aspirations.

5.3 Policy implications and future research

The analysis in this thesis uses robust empirical evidence with the aim to inform policy makers about the determinants of educational aspirations and choices, and to guide them towards developing the English educational system, improving the outcomes of disadvantaged students and providing equality of opportunity in education and the labour market. The results of this thesis highlight the importance of background factors in influencing educational aspirations and choices. Further, they establish the positive relationship between academic match in post-compulsory education and labour market income returns. Although the study claims no causality of this finding, the positive direction of this relationship is statistically significant while holding important background factors constant (prior achievements at ages 11, 14 and 16 and demographic characteristics) and while estimating separate models for heterogeneous groups of students (girls vs. boys,

university vs. non-university participants). Such evidence is convincing that the educational choices of young people are significant for their future and that there is no unique 'best' choice for all students, but that such decisions vary across individuals.

There are important policy implications to be drawn from these findings. The results presented in Chapter 2 identified that the educational aspirations and choices of young people, conditioning on individual ability, are largely driven by social class. Although the socio-economic effects have been markedly reduced between the years, they are still present. Further, the chapter identified a decline in vocational education participation which did not arise from falling aspirations but from rising aspirations and actual participation in academic education. The findings of the third empirical study, presented in Chapter 4, established the important role of making the 'right' educational decisions, which would be those matching the students' academic credentials, for labour market income returns and also showed that the probability of making the 'right' decision, being matched to their 16-19 post-compulsory qualification, is also driven by social class.

Finally, the second empirical study of this thesis, presented in Chapter 3, examines the importance of peer influences on educational aspirations. The findings of Chapter 3 have implications for the allocation of students across schools. Even in a mostly comprehensive education system as in England, with no selection by schools on ability, there are still large differences in student intakes across schools, in terms of their socio-economic background and prior ability; mostly associated with better-off families being able to pay higher house prices closer to high-performing schools (Gibbons and Machin, 2003). While in some cases individuals may be inspired to undertake a route that turns out to be beneficial but which they might not otherwise have chosen, in other cases, they may be influenced by their peers to take a less advantageous or appropriate route. For example, some individuals could follow their academically-orientated peers when a vocational course may have been more suitable for them, while other individuals in a vocationally-dominant peer group may be more suited to academic study themselves. In such cases where an individual could be influenced into making the 'wrong' choice for their own personal circumstances by simply following their peers, the analysis of peer effects on pupils who received educational information, advice and guidance suggests that such advice could play an important role given that it is shown to weaken the influence of peers. Given the difficulty of identifying a priori those pupils who would make the 'wrong' choice if following their peers, this suggests the importance of providing educational advice and guidance to all pupils as an alternative

source of information to guide choices.

In the future, as an expansion of Chapter 4, it would be interesting to assess whether increased provision of information, advice and guidance in schools regarding available post-compulsory education routes is efficient in decreasing academic mismatch in 16-19 education. Further, the role of educational and occupational aspirations in influencing academic match would be an interesting area to investigate. Rich individual level survey data, such as that used in Chapters 2 and 3 of this thesis, would be suitable to be used to expand this investigation and study whether the lack of educational advice or the heterogeneous educational and professional aspirations among students are driving some of the socio-economic effect identified on academic match in the analysis of Chapter 4.

Overall, the findings of this thesis establish that educational aspirations and choices are influenced by background factors other than individual ability and that they are, to a large extent, socially graded. The complexity of the vocational system in England is likely to be one major aspect of the socio-economic inequalities identified in educational attainment. The proliferation of vocational qualifications in England has led to a system little understood not only by students but by employers as well. If employers are not even sure what a person has learned as a result of taking a particular vocational qualification, it is unsurprising that some qualifications have very little economic value (Machin and Vignoles, 2018). Given the low returns identified for vocational qualifications, it is not surprising that they became the option of the least able and most disadvantaged individuals.

Efforts should be made to improve the vocational system, making it more accessible and easier to understand, in order to make young individuals aspire and choose to follow it. Policy-makers interested in social mobility should be focusing more on providing students with information related to the available 16-19 education courses that are suitable to each student's ability credentials and future educational and occupational aspirations.

In summary, this thesis contributes to various related strands of literature in the field of the Economics of Education. In its entirety, this thesis builds upon and develops existing academic research on the literature concerned with the determinants of educational decisions and with social mobility. While there are limitations to the research presented in this thesis, several future avenues have been discussed to develop further understanding. The implications drawn from this research should be important for every policy maker, social scientist, teacher and parent interested in social mobility and equality of opportunity.

References

Acemoglu, D. and J.-S. Pischke (1999). Beyond Becker: Training in imperfect labour markets. *Economic Journal 109*, 112–142.

Ammermueller, A. and J.-S. Pischke (2009). Peer effects in European primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics 27*(3), 315–348.

Angrist, J. D. and K. Lang (2004). Does school integration generate peer effects? evidence from Boston's metco program. *American Economic Review 94*(5), 1613–1634.

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

Atkinson, A., S. Burgess, P. Gregg, C. Propper, and S. Proud (2008). The impact of classroom peer groups on pupil GCSE results. *Centre for Market and Public Organisation WP* (No. 08/187).

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

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press.

Belasco, A. S. and M. J. Trivette (2015). Aiming low: Estimating the scope and predictors of postsecondary undermatch. *The Journal of Higher Education 86*(2), 233–263.

Benabou, R. (1996). Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance. *86*(3), 584–609.

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

Betts, J. R. and A. Zau (2004). Peer groups and academic achievement: Panel evidence from administrative data. [Online; Accessed 19-July-2017] Available from: http://legacy.iza.org/iza/en/webcontent/events/transatlantic/papers_2004/betts.pdf.

Black, S. E., K. E. Cortes, and J. A. Lincove (2015). Academic undermatching of high-achieving minority students: Evidence from race-neutral and holistic admissions policies. *American Economic Review 105*(5), 604–10.

Blanden, J. and S. Machin (2007). *Recent changes in intergenerational mobility in Britain*. Sutton Trust London.

Blanden, J. and S. Machin (2010). Intergenerational inequality in early years assessments. *Children of the 21st century–the first five years*, 153–168.

Bobonis, G. J. and F. Finan (2009). Neighborhood peer effects in secondary school enrollment decisions. *Review of Economics and Statistics 91*(4), 695–716.

Boozer, M. A. and S. E. Cacciola (2001). Inside the 'Black Box' of Project STAR: Estimation of Peer Effects using Experimental Data. *Yale Economic Growth Center Discussion Paper* (832).

Bowen, H. P. and M. F. Wiersema (2004). Modeling limited dependent variables: Methods and guidelines for researchers in strategic management. In *Research methodology in strategy and management*, pp. 87–134. West Yorkshire, England: Emerald Group Publishing Limited.

Bradley, S. and J. Taylor (2004). Ethnicity, educational attainment and the transition from school. *Manchester School 72*(3), 317–346.

Brunello, G., M. De Paola, and V. Scoppa (2010). Peer effects in higher education: Does the field of study matter? *Economic Inquiry 48*(3), 621–634.

Burgess, S., B. McConnell, C. Propper, and D. Wilson (2004). Sorting and choice in English secondary schools. *Centre for Market and Public Organisation Working Paper 4*.

Burke, M. A. and T. R. Sass (2013). Classroom peer effects and student achievement. *Journal of Labor Economics 31*(1), 51–82.

Campbell, S., L. Macmillan, R. Murphy, G. Wyness, et al. (2019). Inequalities in student to course match: Evidence from linked administrative data. Technical report, Centre for Economic Performance, LSE.

Carneiro, P., L. Dearden, and A. Vignoles (2010). The economics of vocational education and training. *International Encyclopedia of Education*, 255–261.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics 27*(3), 439–464.

Cheng, S. and J. S. Long (2007). Testing for IIA in the multinomial logit model. *Sociological methods & research 35*(4), 583–600.

Chevalier, A., C. Harmon, V. OSullivan, and I. Walker (2013). The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics 2*(1), 8.

Chevalier, A. and G. Lanot (2002). The Relative Effect of Family Characteristics and Financial Situation on Educational Achievement. *Education Economics 10*(2), 165–181.

Chowdry, H., C. Crawford, L. Dearden, A. Goodman, and A. Vignoles (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.

Cohen-Cole, E. and J. M. Fletcher (2008). Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics 27*(5), 1382–1387.

Coleman, J. S. (1968). Equality of educational opportunity. *Integrated Education 6*(5), 19–28.

Cooley, J. (2014). Can achievement peer effect estimates inform policy? A view from inside the black box. *Review of Economics and Statistics 96*(3), 514–523.

Crawford, C. (2014). Socio-economic differences in university outcomes in the UK: drop-out, degree completion and degree class. *IFS Working Paper W14/31*.

Crawford, C., P. Gregg, L. Macmillan, A. Vignoles, and G. Wyness (2016). Higher education, career opportunities, and intergenerational inequality. *Oxford Review of Economic Policy 32*(4), 553–575.

Croll, P. (2008). Occupational choice, socioeconomic status and educational attainment: a study of the occupational choices and destinations of young people in the British Household Panel Survey. *Research Papers in Education 23*(3), 243–268.

Czepiel, S. A. (2012). Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation. pp. 1–23. [Online; Accessed 21-September-2017] Available from: https://czep.net/stat/mlelr.pdf.

Davies, R., E. Heinesen, and A. Holm (2002). The relative risk aversion hypothesis of educational choice. *Journal of Population Economics 15*(4), 683–713.

De Coulon, A. and C. Greenwood (2015). The 1980s reforms of vocational education in Britain: a longitudinal perspective. *Unpublished manuscript presented at CVER Conference 2016*.

De Fraja, G. (2002). The design of optimal education policies. *Review of Economic Studies 69*(2), 437–466.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2009). Be as careful of the company you keep as of the books you read: peer effects in education and on the labor market. *National Bureau of Economic Research* (Discussion Paper No. w14948).

De Groote, O. (2017). The effects of high school curriculum: A model of program and effort choice. *Conference paper*. [Online; Accessed 25-August-2018] Available from: http://pareto.uab.es/pdf/seminars/jobmarket/2018/Olivier%20De%20Groote.pdf.

de Xavier Pinto, C. C. (2010). Semiparametric estimation of peer effects in classrooms: Evidence for Brazilian schools in 2003. [Online; Accessed 02-October-2016] Available from: https://www.economics.uci.edu/files/docs/econoseminar/w10/pinto.pdf.

Dearden, L., S. Machin, and A. Vignoles (2009). Economics of education research: a review and future prospects. *Oxford Review of Education 35*(5), 617–632.

Delebarre Jeanne (2016). Apprenticeship Statistics: England (1996-2015). *House of Commons library* (06113), 1–15.

Di Stasio, V., T. Bol, and H. G. Van de Werfhorst (2015). What makes education positional? Institutions, overeducation and the competition for jobs. *Research in Social Stratification and Mobility 43*, 53–63.

Dillon, E. W. and J. A. Smith (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics 35*(1), 45–66.

Dills, A. K. (2005). Does cream-skimming curdle the milk? A study of peer effects. *Economics of Education Review 24*(1), 19–28.

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.

Dilnot, C. (2018). The relationship between A-level subject choice and league table score of university attended: the facilitating, the less suitable, and the counter-intuitive. *Oxford Review of Education 44*(1), 118–137.

Ding, W. and S. F. Lehrer (2007). Do peers affect student achievement in China's secondary schools? *Review of Economics and Statistics 89*(2), 300–312.

Dodgeon, B. and S. Parsons (2011). Deriving Highest Qualification in NCDS and BCS70. *Centre for Longitudinal Studies*, 0–39. [Online; Accessed 14-March-2016] Available from: https://cls.ucl.ac.uk/wp-content/uploads/2018/06/Deriving-highest-qualification-in-NCDS-and-BCS70.pdf.

Dolton, P. and A. Vignoles (2000). The incidence and effects of overeducation in the U.K. graduate labour market. *Economics of Education Review 19*(2), 179–198.

Egerton, M. and A. H. Halsey (1993). Trends by Social Class and Gender in Access to Higher Education in Britain. *Oxford Review of Education 19*(2), 183–196.

Erberber, E., M. Stephens, S. Mamedova, S. Ferguson, and T. Kroeger (2015). Socioeconomically disadvantaged students who are academically successful: Examining academic resilience cross-nationally. policy brief no. 5. *International Association for the Evaluation of Educational Achievement*.

Ermisch, J. and M. Francesconi (2002). The Effect of Parents' Employment on Children's Educational Attainment. *ISER Working Papers Number 2002-21* (215).

Evans, W. N., W. E. Oates, and R. M. Schwab (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy 100*(5), 966–991.

Facchinello, L. (2017). Does peer ability affect education choices? *SSRN*. [Online; Accessed 23-September-2017] Available from: http://dx.doi.org/10.2139/ssrn.2966549.

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

Fernandez, R. and R. Rogerson (1996). Income distribution, communities, and the quality of public education. *The Quarterly Journal of Economics 111*(1), 135–164.

Fernandez, R. and R. Rogerson (1998). Public education and income distribution: A dynamic quantitative evaluation of education-finance reform. *American Economic Review 88*(4), 813–833.

Fertig, M. (2003). Educational production, endogenous peer group formation and class composition-evidence from the PISA 2000 study. *IZA Discussion Paper Series 2003, No. 714*.

Foster, G. (2003). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics 90*(8), 1455–1475.

Fry, T. R. and M. N. Harris (1996). A monte carlo study of tests for the independence of irrelevant alternatives property. *Transportation Research Part B: Methodological 30*(1), 19–30.

Fry, T. R. and M. N. Harris (1998). Testing for independence of irrelevant alternatives some empirical results. *Sociological Methods & Research 26*(3), 401–423.

Galindo-Rueda, F., O. Marcenaro-Gutierrez, and A. Vignoles (2004). The Widening Socio-Economic Gap in UK Higher Education. *National Institute Economic Review* (190), 75–88.

Galindo Rueda, F. and A. Vignoles (2003, December). Class Ridden or Meritocratic? An Economic Analysis of Recent Changes in Britain. *IZA Discussion Paper* (677).

Galindo-Rueda, F. and A. Vignoles (2005a). The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability. *Centre for the Economics of Education DP 52*.

Galindo-Rueda, F. and A. Vignoles (2005b). The Declining Relative Importance of Ability in Predicting Educational Attainment. *Journal of Human Resources 40*(2), 335–353.

Gaviria, A. and S. Raphael (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics 83*(2), 257–268.

George, T. and M. Jr. (1996, June). Money, expectations and the U.S. civil war. *American Economic Review 86*(3), 661–671.

Gesthuizen, M. and P. Scheepers (2010). Economic Vulnerability among Low-Educated Europeans: Resource, Composition, Labour Market and Welfare State Influences. *Acta Sociologica 53*(3), 247–267.

Gibbons, S. and S. Machin (2003). Valuing English primary schools. *Journal of Urban Economics 52*(2), 197–219.

Gibbons, S., O. Silva, and F. Weinhardt (2013). Everybody needs good neighbours? Evidence from students outcomes in England. *Economic Journal 123*(571), 831–874.

Gibbons, S. and S. Telhaj (2016). Peer effects: Evidence from secondary school transition in England. *Oxford Bulletin of Economics and Statistics 78*(4), 548–575.

Glennerster, H. (1991). Quasi-markets for education? *Economic Journal 101*(408), 1268–1276.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

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

Goux, D. and E. Maurin (2007). Close neighbours matter: Neighbourhood effects on early performance at school. *Economic Journal 117*(523), 1193–1215.

Graham, B. S. (2008). Identifying Social Interactions Through Conditional Variance Restrictions. *Econometrica 76*(3), 643–660.

Gutman, L. and R. Akerman (2008). Determinants of aspirations. *Centre for Research on the Wider Benefits of Learning Research Report No. 27, Institute of Education, University of London*.

Guyon, N. and E. Huillery (2016). Biased aspirations and social inequality at school: Evidence from french teenagers: Evidence from french teenagers. Technical report. [Online; Accessed 15-June-2017] Available from: http://econ.sciences-po.fr/sites/default/files/file/elise/paper_AspirationsFailures_EJ_R1.pdf.

Hanushek, E., J. Kain, J. Markman, and S. Rivkin (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics 18*(5), 527–544.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis 21*(2), 143–163.

Hausman, J. and D. McFadden (1984). Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society 52*, 1219–1240.

Hearn, J. C. (1991). Academic and non-academic influences on the college destinations of 1980 high school graduates. *Sociology of Education 64*(3), 158–171.

Hibbert, T. (2010). Face to face: what young people think about information, advice and guidance. *National Youth Agency*. [Online; Accessed 19-July-2017] Available from: https://www.lgcplus.com/Journals/3/Files/2010/9/7/NYA%20Face-to-Face%20IAG%20report.pdf.

Hoel, J., J. Parker, and J. Rivenburg (2005). Peer Effects: Do First-Year Classmates, Roommates, and Dormmates Affect Students' Academic Success? *Higher education data sharing consortium winter conference, Santa Fe, NM*.

Holm, A. and M. M. Jaeger (2008). Does relative risk aversion explain educational inequality? a dynamic choice approach. *Research in Social Stratification and Mobility 26*(3), 199–219.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. *National Bureau of Economic Research, Cambridge, MA* (Working Paper no.7867).

Hoxby, C. M. and C. Avery (2012). The missing "one-offs": The hidden supply of high-achieving, low income students. Technical report, National Bureau of Economic Research.

Hoxby, C. M. and G. Weingarth (2006). Taking race out of the equation: School reassignment and the structure of peer effects. [Online; Accessed 20-July-2017] Available from: https://www.pausd.org/sites/default/files/pdf-faqs/attachments/TakingRaceOutOfTheEquation.pdf.

Hupkau, C., S. McNally, J. Ruiz-Valenzuela, and G. Ventura (2017). Post-compulsory education in England: choices and implications. *National Institute Economic Review 240*(1), R42–R57.

Jenkins, A., A. Vignoles, A. Wolf, and F. Galindo-Rueda (2003). The determinants and labour market effects of lifelong learning. *Applied Economics 35*(16), 1711–1721.

Jenkins, H., K.-L. English, O. Hristova, A. Blankertz, V. Pham, and C. Wilson (2017). Social mobility and economic success: How social mobility boosts the economy. *Sutton Trust*. [Online; Accessed 13-May-2019] Available from: https://www.suttontrust.com/wp-content/uploads/2017/07/Oxera-report_WEB_FINAL.pdf.

Jerrim, J. and A. Vignoles (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.

Jonsson, J. O. and C. Mood (2008). Choice by contrast in Swedish schools: How peers' achievement affects educational choice. *Social Forces 87*(2), 741–765.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics 61*(3), 458–495.

Keep, E. and K. Mayhew (2004). The economic and distributional implications of current policies on higher education. *Oxford Review of Economic Policy 20*(2), 298–314.

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

Lavy, V., M. D. Paserman, and A. Schlosser (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *Economic Journal 122*(559), 208–237.

Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of Urban Economics 56*(2), 169–191.

Lin, X. (2010). Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics 28*(4), 825–860.

Lyle, D. S. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point. *The Review of Economics and Statistics 89(2)*, 289–299.

Lynch, L. M. (1993). The Economics of Youth Training in the United States. *Economic Journal 103*(420), 1292–1302.

Machin, S. and A. Vignoles (2018). *What's the Good of Education? The Economics of Education in the UK*. Princeton: Princeton University Press.

Madhu, B., N. C. Ashok, and S. Balasubramanian (2014). A Multinomial Logistic Regression Analysis to Study The Influence Of Residence And Socio-Economic Status On Breast Cancer Incidences In Southern Karnataka. *International Journal of Mathematics and Statistics Invention 2*(5), 1–8.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies 60*(3), 531–542.

Marsh, H. W. and K.-T. Hau (2003). Big-fish–little-pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist 58*(5), 364.

Mayer, S. E. and C. Jencks (1989). Growing up in poor neighborhoods: How much does it matter? *Science 243*(4897), 1441–1445.

McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review 22*(2), 131–141.

McEwan, P. J. and K. A. Soderberg (2006). Roommate effects on grades: Evidence from first-year housing assignments. *Research in Higher Education 47*(3), 347–370.

McIntosh, S. (2006). Further analysis of the returns to academic and vocational qualifications. *Oxford Bulletin of Economics and Statistics 68*(2), 225–251.

McIntosh, S. and D. Morris (2016). Labour market returns to vocational qualifications in the labour force survey. *CVER Research Paper 6*.

McIntosh, S. and D. Morris (2018). Labour market outcomes of older versus younger apprentices: A comparison of earnings differentials. *CVER Discussion Paper series DP016*.

Mendolia, S., A. R. Paloyo, and I. Walker (2018). Heterogeneous effects of high school peers on educational outcomes. *Oxford Economic Papers 70*(3), 613–634.

Moffitt, R. A. et al. (2001). Policy interventions, low-level equilibria, and social interactions. *Social Dynamics 4*(45-82), 6–17.

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

Parsons, S. (2014). Childhood cognition in the 1970 british cohort study. *CLS Data Note*.

Parsons, S., I. Schoon, R. Rush, and J. Law (2011). Long-term outcomes for children with early language problems: Beating the odds. *Children & Society 25*(3), 202–214.

Payne, J. (2003). Vocational Pathways at Age 16-19. *Department for Education and Skills Report No. 501*.

Rivkin, S. G. (2001). Tiebout sorting, aggregation and the estimation of peer group effects. *Economics of Education Review 20*(3), 201–209.

Robertson, D. and J. Symons (2003). Do peer groups matter? Peer group versus schooling effects on academic attainment. *Economica 70*(277), 31–53.

Roderick, M., V. Coca, and J. Nagaoka (2011). Potholes on the road to college: High school effects in shaping urban students participation in college application, four-year college enrollment, and college match. *Sociology of Education 84*(3), 178–211.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Sacerdote, B. (2011). Peer effects in education: How might they work, are they and how much do we know thus far? *Handbook of the Economics of Education 3*, 249–277.

Sallee, J. M., A. M. Resch, and P. N. Courant (2008). On the optimal allocation of students and resources in a system of higher education. *The BE Journal of Economic Analysis & Policy 8*(1), 1–26.

Sanbonmatsu, L., J. R. Kling, G. J. Duncan, and J. Brooks-Gunn (2005). Neighborhoods and academic achievement results from the Moving to Opportunity experiment. *Journal of Human Resources 41*(4), 649–691.

Schmidheiny, K. (2007). Multinomial Choice (Basic Models). [Online; Accessed 30-November-2015] Available from: https://www.schmidheiny.name/teaching/multinomialchoice2up.pdf.

Schoon, I. (2006). *Risk and resilience: Adaptations in changing times*. Cambridge University Press.

Schoon, I. and K. Duckworth (2010). Leaving school early–and making it! *European Psychologist 15*(4), 283–292.

Scott, J. and G. Marshall (2009). *A Dictionary of Sociology*. Oxford Dictionary of Sociology. Oxford: Oxford University Press.

Scott Long, J. (1997). *Regression models for categorical and limited dependent variables*, Volume 7. Thousand Oaks, CA, US: Sage Publications, Inc.

Small, K. A. and C. Hsiao (1985). Multinomial logit specification tests. *International Economic Review 26*(3), 619–627.

Smith, J., M. Pender, and J. Howell (2013). The full extent of student-college academic undermatch. *Economics of Education Review 32*, 247–261.

Social Mobility Commission (2019). State of the nation 2018-2019: Social mobility in great britain. [Online; Accessed 13-May-2019] Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/798687/SMC_State_of_Nation_2018-19_Summary.pdf.

Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from Project STAR. *Economic Journal 123*(569), 574–605.

Speckesser, S. and S. Hedges (2017). Peer effects and social influence in post-16 educational choice. *National Institute of Economic and Social Research, Discussion Paper No. 483*.

Standards and Testing Agency (2012). *Key stage 2 mathematics test framework*. [Online; Accessed 25-November-2015] Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/628819/2016_KS2_Mathematics_framework_PDFA_V5.pdf.

Stinebrickner, R. and T. R. Stinebrickner (2003). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics 90*(8), 1435–1454.

Van de Werfhorst, H. G. (2011). Skills, positional good or social closure? The role of education across structural–institutional labour market settings. *Journal of Education and Work 24*(5), 521–548.

Vigdor, J. and T. Nechyba (2007). Peer effects in North Carolina public schools. *Schools and the equal opportunity problem CESifo*, 73–101.

Vignoles, A., A. De Coulon, and O. Marcenaro-Gutierrez (2011). The value of basic skills in the British labour market. *Oxford Economic Papers 63*(1), 27–48.

Von Hinke, S., G. Leckie, and C. Nicoletti (2019). The use of instrumental variables in peer effects models. *Oxford Bulletin of Economics and Statistics 10.1111/obes.12299*.

West, A. and A. Hind (2003). Secondary School Admissions in England: Exploring the Extent of Overt and Covert Selection. *Report for Research and Information on State Education Trust*.

Winston, G. and D. Zimmerman (2004). *Peer effects in Higher Education*. Chicago: University of Chicago Press.

Wolf, A. (2011). Review of vocational education: The Wolf report. *Department for Education and Department for Business Innovation and Skills*. [Online; Accessed 09-April-2019] Available from: https://dera.ioe.ac.uk/11621/1/DFE-00031-2011.pdf.

Woodhall, M. (1967). The economics of education. *Educational Organization, Administration and Finance 37*(4), 387–398.

Wulff, J. N. (2015). Interpreting results from the multinomial logit model demonstrated by foreign market entry. *Organizational Research Methods 18*(2), 300–325.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.