National Institute of Economic and Social Research

PEER EFFECTS AND SOCIAL INFLUENCE IN POST-16 EDUCATIONAL CHOICE

Sophie Hedges[1,2]
Stefan Speckesser[1,3]

[1] Centre for Vocational Education Research
[2] London Economics
[3] National Institute of Economic and Social Research

NIESR Discussion Paper No. 483

Date: 8 December 2017

National Institute of Economic and Social Research

About the National Institute of Economic and Social Research

The National Institute of Economic and Social Research is Britain's longest established independent research institute, founded in 1938. The vision of our founders was to carry out research to improve understanding of the economic and social forces that affect people's lives, and the ways in which policy can bring about change. Seventy-five years later, this remains central to NIESR's ethos. We continue to apply our expertise in both quantitative and qualitative methods and our understanding of economic and social issues to current debates and to influence policy. The Institute is independent of all party political interests.

National Institute of Economic and Social Research
2 Dean Trench St
London SW1P 3HE
T: +44 (0)20 7222 7665
E: enquiries@niesr.ac.uk
niesr.ac.uk
Registered charity no. 306083

This paper was first published in December 2017

Peer Effects and Social Influence in Post-16 Educational Choice

Sophie Hedges and Stefan Speckesser

Abstract

This paper investigates whether the educational choices that young people make after the completion of their GCSEs (at age 16) are influenced by their peers. More specifically, it takes advantage of the variation in peer groups that arises when students move from primary to secondary school in order to isolate the impact of secondary school peers on the choice of educational trajectory. These trajectories are broadly classified as academic, vocational, a combination of the two, or no education at all. In order to overcome the common problems associated with the identification of peer effects, the ability of the primary school peers of secondary school peers, who are not going to the same secondary school, is used as an instrument for secondary school peer group quality. These 'peers of peers' did not go to the same primary or secondary school as the individual of interest and so cannot have had any direct impact on them. Our results show that higher ability peers reduce the likelihood that an individual will choose a vocational course at age 16 after controlling for the individual's own ability.

We also find a very strong effect of household income on education choices, showing that the more deprived a student's background is, the more likely they are to opt for a vocational trajectory over an academic one.

Acknowledgements

We would like to thank Andy Dickerson, Richard Dickens, Matthew Bursnall and Sandra McNally for helpful comments and assistance with earlier versions of this paper, as well as other CVER members for attending an internal presentation of the work in progress. We would also like to thank the Department for Education (DfE) for their help with access to the data used in this project.

This paper was originally published by the Centre for Vocational Education Research (CVER) as Research Paper 008 in November 2017.

Contact details

Stefan Speckesser (s.speckesser@niesr.ac.uk), National Institute of Economic and Social Research, 2 Dean Trench Street, London SW1P 3HE

1 Introduction

Vocational education is commonly assumed to be a route for lower ability students who are not able to achieve an A level qualification. Clearly ability is an important determinant, given that most A level courses generally require a high level of GCSE achievement as a prerequisite, but there remain students with strong exam results who choose to pursue a vocational route once they have completed compulsory schooling. Furthermore, it is not necessarily the case that the pupils following a vocational trajectory are veering away from pursuing education at a high level; although it is less common than for A level students, there are a significant number of individuals who proceed into higher education after achieving vocational qualifications. Finally, there are individuals who enrol in qualifications only to drop out of the course before completion. These individuals are all considered to make educational choices which do not reflect their revealed ability, and currently there is little known about the factors which influence this, supposedly irrational, choice.

It has been well documented in the social science literature that peers can have an influence on individual behaviour and choices in a variety of diverse settings. These include, but are not limited to, the consumption of drugs/alcohol (Gaviria & Raphael 2001), risk taking (Lahno & Serra-Garcia 2015), productivity (Mas & Moretti 2009), obesity (Trogdon et al. 2008), and criminal behaviour (Bayer et al. 2009). It is not inconceivable, therefore, that social influence may account for some of the inconsistencies in decisions made regarding educational trajectories. Such a finding, if this turns out to be the case, would fill some of the gap in the currently unknown determinants of post-16 educational choice. Furthermore, there could be very real policy implications arising from this research; a proper understanding of the choice process is crucial for the provision of appropriate information and guidance for young people about the available schooling options and the consequences of their decisions.

This paper will investigate the role of peer effects on post-compulsory educational choice in terms of whether individuals opt for an academic route, a vocational route, a combination of the two, or no educational choice at all. The empirical strategy is based on that proposed by Mendolia et al. (2016), who used peers of peers to isolate peer effects on academic outcomes at secondary school.

2 Background and existing literature

2.1 Theoretical background

Peer effects refer to the influence of a group on an individual and arise from the observation that group members are inclined to behave in a similar manner. Manski (1993) provided three hypotheses to explain this; individuals may behave in similar ways because they have similar characteristics (correlated effects), their behaviour may vary with the behaviour of the group (endogenous effects) or their behaviour may vary with the exogenous characteristics of the group (exogenous effects). To the extent that there are endogenous effects, peer effects are essentially an externality (or spillover) of human capital from one individual to another.

Any study conducted in the area of peer effects must encounter and attempt to overcome two common problems. Firstly, individuals typically endogenously sort into groups. This makes it difficult to disentangle correlated effects from endogenous effects. In other

words, if an individual behaves in a way that is consistent with that of the group, then their behaviour may be incorrectly assumed to have been influenced by the group when, in actual fact, it may be the case that similar people who already behave in similar ways have grouped themselves together. The second issue is that all members of the group influence each other at the same time so that is not clear whether the actions and attributes of an individual's peers influence them or imitate them (Manski referred to this concept as the reflection problem). This can best be explained by considering two individuals, X and Y, who have exactly the same friends so that they are each a member of each other's peer groups. Hence, as the actions and choices of X affect Y, Y affects X simultaneously. This is true of all members of the peer group. Consequently, it is impossible to isolate the influence that X has on the peer group as everyone else will exert an influence on the peer group, including X, at the same time.

The endogenous sorting problem can easily be overcome via the random assignment of peer groups (although it is not necessarily easy to find examples of this in reality). In order to combat the reflection problem, an exogenous source of variation in peer groups, in which individuals do not sort themselves, is required. This then allows for the isolation of the influence of a particular group composition on the individual. Any measures of the characteristics of group members, for example an ability proxy such as exam results, should be measured prior to the peer group formation in order to eliminate any possibility that these characteristics have already been influenced by the group at the time that they were recorded.

2.2 Existing literature

Much of the existing literature in the field of education economics focuses on the influence of peers on individuals' educational achievement (Hanushek et al. 2003, Kiss 2013, Vardardottir 2013), but more recent studies are now looking into the effect that peers can have on academic choice. For example, Poldin et al. (2015) find that specialisation choice among Russian undergraduates is strongly influenced by friends as well as study partners and Ashworth & Evans (2001) suggest that, for females, the decision to study economics depends on the proportion of females studying that subject. This paper will contribute to the emerging literature on the effects of peers on academic choice, but will provide a novel contribution by investigating not just subject specialisation, but whether peers can influence the decision to pursue an academic or vocational track after the completion compulsory schooling.

There have been a number of approaches to combat the problems commonly associated with the identification of peer effects. Early studies have utilised the random assignment of roommates in college dormitories to overcome any endogenous sorting (Sacerdote 2001, Stinebrickner & Stinebrickner 2006, Zimmerman 2003). In the most well-known of these studies, Sacerdote (2001) found that freshman roommates at Dartford College were significant for Grade Point Averages (GPAs) and participating in societies such as fraternities, but not for the choice of college major. However, it is not necessarily the case that college roommates spend much time together, which would weaken the potential influence that they could have on each other, and so this cannot necessarily be taken as evidence that peers do not affect college major at all.

A second option for overcoming some of these issues is to exploit a natural experiment. For example, Cipollone & Rosolia (2007) took advantage of the random occurrence of a 1980 Italian earthquake which exempted men from certain towns from compulsory

military service, and found that the exemption resulted in increased secondary school graduation rates. In an alternative approach, Carrell et al. (2009) used a dataset where students were randomly assigned to peer groups at the United States Air Force Academy (USAFA), and they were required to spend the majority of their time with this group. The study found large effects on academic achievement which persisted in the following years. Random, but non-natural, experiments have also been used; Graham (2008) found a significant effect of peers in their study of the classroom reduction programme which was implemented in the Tennessee Project STAR.

Social network analysis is increasingly being utilised as a method of developing a more accurate picture of an individual's closest peers and overcoming some of the identification problems typically associated with the estimation of peer effects (Bramoullé et al. 2009, Calvó-Armengol et al. 2009). Calvó-Armengol et al. (2009) argue that it may not be the peers themselves that matter, but the individual's position within their friendship network (their 'centrality'), and that increases in centrality are associated with better academic performance.

De Giorgi et al. (2010) used data from an Italian university for a cohort of students who faced a common first year before choosing a major of either business or economics. They then used the composition of classes in the first year to identify partially overlapping peer groups, utilising the excluded peers of these groups (i.e. the friends of friends) as an exclusion restriction, and investigated the influence of peers on choice of major. They found that people are more likely to choose the major that the majority of their peers choose, even if they had a comparative ability advantage in the alternative subject. This ultimately led to worse academic performance and subsequent wages and job satisfaction.

In one of the few studies based on UK data, Gibbons & Telhaj (2015) and Lavy et al. (2012), who both used the National Pupil Database (NPD), defined the peer group to consist of all individuals in a given year group at a given institution. In order to allow for identification, they utilised the fact that, in the UK, most children change schools (from primary to secondary) at age 11 and that the majority of students experience a huge change in their peer group as a result of this change (when they reach secondary school, 88% of a students' cohort consists of new peers on average). As they were able to identify other students who made the same primary-to-secondary school transition, they could account for the influence of the prior ability of new peers (measured as achievement at age 11 in Key Stage 2 tests taken at primary school before the transition) on academic achievement in national tests taken at age 14. They both found a small but significant effect of peers, with low ability peers being more influential than those of average or high ability in the case of the latter paper. However, Burke & Sass (2013) and Carrell et al. (2009) found that classroom-level peers had much more of an effect on individual achievement than when the entire year group was considered, so these results may underestimate the effect of those that individuals interact with a lot.

Recently, Mendolia et al. (2016) have bridged the gap between the De Giorgi et al. (2010) paper and the Gibbons & Telhaj (2015) and Lavy et al. (2012) papers. They used the Longitudinal Study of Young People in England (LSYPE), and similarly took advantage of the UK primary to secondary school transition, but they employed a peers of peers methodology. Specifically, they used the primary school peers of secondary school peers, who did not go to the same primary school as the individual of interest, as an instrument for secondary school peer group ability. They found evidence of peer effects for test scores, particularly at the lower end of the distribution, and some indication that peer ability could also influence the chance that students take A levels.

Given that the decision to pursue a vocational or academic track is taken at age 16, when students typically still live at home (ignoring boarding schools), it will not be possible to apply any analysis utilising a roommates set up. Additionally, in the absence of a natural experiment or detailed data on social interactions or classroom compositions, the choice of remaining identification strategies are limited. This paper proposes to utilise the peers of peers approach implemented by Mendolia et al. (2016) in order to identify peer effects, but to apply this methodology to isolate the influence of peers on the decision to invest in either academic or vocational education at age 16.

2.3 Institutional context

In the UK, there are five 'Key Stages' of education. Students complete Key Stages 1 and 2 during primary school, progressing to secondary school after the completion of Key Stage 2 at age 11. In secondary school, Key Stage 3 is taken up until age 14 (although national assessment of this phase was abolished in 2008) after which Key Stage 4 is taken for two years, culminating in GCSE examinations at age 16. Education is compulsory up until this point, after which students can either pursue an academic route (e.g. A levels/International Baccalaureate Diploma), a vocational route (e.g. BTECs/NVQs), a combination of the two, or leave education altogether. Recently, the compulsory participation age rose to 18, but students still face a choice between academic and vocational qualifications at age 16.

When students move from primary to secondary schools they face a choice which is constrained by the availability of suitable schools within a given area (i.e. a commutable distance from their home). Parents apply to the Local Authority (LA) with an ordered list of their preferred schools, and the LA then allocates students to schools based on criteria which vary in each area. For example, preferences may be given to students who already have a sibling at the school, live close to the school, or have special educational needs[1]. The outcomes of the application process are strongly dependent on other factors which vary from year to year. For example, if a given year has more applicants than in the year before, or there are more pupils with special education needs, then a given student may be allocated to a different school based purely on the year in which they applied. The implication is that there is a random component to the application process and, hence, the ability and background characteristics of a peer group will randomly vary from year to year.

Another important element of the UK's education sector is that it is typically not the case that students are taught in one tutor group for the duration of their time at secondary school; the teaching groups are composed according to the subjects chosen for GCSEs and ability setting or other timetabling requirements. This feature makes it increasingly likely that the pupils will encounter the majority of the other students in their cohort at some point in their secondary school education, and so it is not necessarily the case that peer groups should be defined at the classroom level.

[1] https://www.gov.uk/schools-admissions/admissions-criteria

3 Empirical analysis

3.1 Identification strategy

This paper will follow the identification method suggested by Mendolia et al. (2016), i.e. it will use the ability of the primary school peers of secondary school peers as an instrument for secondary school peer group quality and apply it to the case of post-16 educational choice. These 'peers of peers' did not go to the same primary or secondary school as the individual of interest and so cannot have had any direct impact on them.

The initial specification originates from the equation below as follows:

$$y_{ips} = \rho a_{ips} + \beta \bar{a}_s + \gamma \mathbf{X_i} + u_i \tag{1}$$

where y_{ips} is the academic choice of individual i who went to primary school p and secondary school s. Prior ability is represented by a_{ips} and $\bar{a}_s$ is the average prior ability and background characteristics of the secondary school peer group so that β is the coefficient of interest. Finally, $\mathbf{X_i}$ is a vector of controls. However, estimates of β in this specification will clearly be endogenous. In order to account for this, peers of peers are used as an instrumental variable. The first stage equation is:

$$\bar{a}_s = \delta \bar{a}_{jh} + \pi \mathbf{X_i} + v_k \tag{2}$$

where $h \neq s$. Here, average secondary school ability $\bar{a}_s$ depends on $\bar{a}_{jh}$, which is the average ability of all of the students who went to the same primary school (j) as those in the secondary school peer group but then progressed to different secondary schools (h). The crucial point is that the peers of peers have not been in either the same primary school or the same secondary school as the individual of interest.

Any sorting driven by selection into schools is overcome by the nature of the primary to secondary school transition in the UK. On average, only six to seven students make the same transition from a given primary school to a given secondary school, moving into secondary school cohorts of roughly 170 students (see below for descriptive statistics). Hence it is clear that there must be a large number of primary schools feeding into each secondary school (Mendolia et al. (2016) claim that more than eight primary schools flow into the vast majority of secondary schools). Given this, peers of peers are likely to have come from an area with different characteristics, particularly given that they are now attending different secondary schools.

3.2 Data

3.2.1 Description of census data used

This paper makes use of data at census level for the entire cohort of secondary school leavers in the summer of 2011, which have been drawn from multiple sources. The National Pupil Database, supplemented by the School Census, provides information about students and their academic achievements in English schools from the age of seven until 18 (Key Stages 1 to Stage 5). Hence the required prior information regarding Key Stage 2 assessments and providers, as well as secondary school providers and assessment outcomes can be obtained from this data. For those who pursue an academic route, their Key Stage 5 decision will also appear in the NPD.

Information for the majority of those who undertook a vocational qualification, or for those who took A Levels at a further education provider as opposed to a sixth form at a school, will be taken from the Individualised Learner Record (ILR), an administrative dataset which contains information on learner participation and achievement in the Further Education sector. This data is supplemented by the Learning Aims Database (LAD), which includes detailed information about the course characteristics, such as the modules undertaken and the number of hours required - Guided Learning Hours (GLH).

A more extensive description of the data processing can be found in the Appendix.

3.2.2 Selection of GCSE cohort

Some limitations will be applied to the sample. Firstly, given that students are required to obtain a certain performance at Key Stage 4 in order to be eligible to undertake A levels, the choice is not really relevant for those who do not achieve this benchmark. In other words, they do not really have a choice to make as the alternative option in this specification is not available to them. Thus the sample will be restricted to look at individuals who achieve at least five GCSEs at grades A*-C (including English and mathematics), which is the commonly accepted pass mark[2].

Secondly, schools which systematically select students on the basis of achievement will also be omitted as these schools are subject to the endogenous sorting problem which compromises the identification strategy. Finally this analysis is concerned with the immediate educational choice, as later choices cannot be reasonable attributed to the secondary school peer groups, and so students who do not enrol in a qualification in the September following the completion of their GCSEs will be considered to have made no educational choice even if they do pursue a course at some later date.

3.2.3 Summary statistics

Our resulting GCSE leaver cohort contains over 400,000 students. There are roughly 15,000 primary schools and approximately 2,600[3] secondary schools. There are, on average, 170 students in a secondary school year group and the mean transition group size (those who go from the same primary to the same secondary school) is only six to seven students.

Table 1 displays the key characteristics of the sample overall, as well as just for those who achieved the pass benchmark of least five GCSEs at A*-C (subsequently denoted GCSE achievers). Roughly half of the sample is female (49.1%), the majority of the students are white (82.7%), and just over half (53.3%) of the students achieved at least five GCSEs at grades A*-C (including English and maths). When looking only at GCSE achievers, the proportion of females increases slightly (to 53.1%), as does the proportion of Asians (from 7.2% to 7.9%), whilst the proportion of other ethnicities falls a little (6.0% to 5.2%).

The post-GCSE educational choice is displayed in Table 2. A description of how the courses were classified into academic or vocational categories is presented in the Appendix (Table 5).

[2]Whilst the choice will only be considered for GCSE achievers, the peer group measure will include all students regardless of achievement.

[3]Only comprehensive secondary schools with at least ten students were retained in order to construct a reasonable peer group measure.

Table 1: Individual Characteristics

	Full Sample		GCSE Achievers	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.491	0.500	0.531	0.499
Ethnicity - White	0.827	0.378	0.830	0.376
Ethnicity - Black	0.041	0.199	0.039	0.194
Ethnicity - Asian	0.072	0.258	0.079	0.269
Ethnicity - Other	0.060	0.237	0.052	0.223
Average Score in Percentiles	48.528	28.726	66.331	22.132
5 GCSEs A*-C (E & M)	0.533	0.499		
Observations	432,366		230,563	

Data: National Pupil Database (NPD). All variables are dummy variables.

Table 2: Educational Choices

	Full Sample		GCSE Achievers	
	Mean	Std. Dev.	Mean	Std. Dev.
Academic only	0.401	0.490	0.679	0.467
Vocational only	0.391	0.488	0.182	0.386
Other	0.008	0.090	0.001	0.035
Academic and vocational	0.050	0.218	0.074	0.262
Any other combination	0.001	0.032	0.001	0.028
None	0.148	0.355	0.063	0.243
Observations	432,366		230,563	

Data: National Pupil Database (NPD). All variables are dummy variables.

Academic qualifications are the most common post-16 choice (40.1%), but vocational qualifications are not that far behind (39.1%). However, when looking at the summary statistics just for GCSE achievers, academic qualifications become much more prominent (67.9%). Interestingly, 18.1% still choose to follow vocational courses, which is evidence against the common misconception that vocational routes are only for low achieving students. Additionally, GCSE achievers are more likely to undertake a combination of academic and vocational courses (7.4%, as opposed to 5.0% for the full sample). Finally, the proportion of students who do not choose education at all is lower when looking solely at GCSE achievers (6.3% compared to 14.8% for the sample overall).

Table 3 shows KS2 Achievement (age 10-11), measured prior to the transition to secondary school. This information will be used as a proxy for peer group quality, as well as to control for an individual's own ability. National Curriculum Level results were used to ensure comparability across schools. These tests include separate papers in English, maths and science. The final measure is the average of all three subjects, after z scores have been calculated, and expressed in terms of percentiles of the total student distribution.

The average score in the maths exam (64%) was higher than for English (59%) and science (57%). As expected, restricting the sample to only those who ultimately went on to achieve five GCSEs at levels A*-C (including English and maths) resulted in higher

Table 3: KS2 Achievement

	Full Sample		GCSE Achievers	
	Mean	Std. Dev.	Mean	Std. Dev.
Raw Points:				
English Score	58.825	16.813	67.981	12.124
Maths Score	64.026	21.717	76.053	15.033
Science Score	57.440	13.155	64.473	8.768
Z scores:				
English Score	-0.042	1.005	0.505	0.725
Maths Score	-0.036	0.997	0.516	0.690
Science Score	-0.070	1.016	0.473	0.677
Composite Measures:				
Average Score	-0.068	0.933	0.497	0.596
Average Score in Percentiles	48.528	28.726	66.331	22.132
Observations	426,872		230,113	

Data: National Pupil Database (NPD). Raw points are total marks achieved in the given test (from 0 to 100). Average score is the average of the English, maths and science points after they have been converted into z scores. Percentiles is the average score expressed in percentiles of the student distribution.

average test scores in all subjects, and lower standard deviations. The final measure to be used in the analysis (average score in percentiles) has a mean of 48.5 and a standard deviation of 28.7 by construction.

3.3 Findings on peers effects and social influence

Table 4 presents the results of the OLS and IV regressions for GCSE achievers, with the OLS estimates serving only as a comparison. In columns (1) and (2), the dependent variable is a dummy variable equal to 1 if the individual chose a vocational course, and 0 if they did not. This is potentially problematic, as the '0' category encapsulates individuals who chose academic courses but also those who left education altogether and entered the labour market, or those who became NEET. Consequently, columns (3) and (4) show the same regressions restricting the sample to just those who enrolled in either a vocational or academic course in the September following the completion of their GCSEs.

The results from column (2), which instruments secondary school peers with 'peers of peers', indicate that a one percentile increase in the measure of individual ability reduces the likelihood that they will enrol in a vocational qualification by 0.34 percentage points on average, holding everything else constant. There is also evidence of peer influence: a one percentile increase in the measure of peer ability reduces the likelihood that an individual will select a vocational programme by 0.2 percentage points on average. This is broadly consistent with the findings of Mendolia et al. (2016)[4]. When the sample is restricted in column (4), the magnitude of these coefficients strengthen slightly and remain highly significant.

The Index of Multiple Deprivation (IMD) was used in place of the traditional Free School Meals (FSM) variable as it provides a more detailed description of the deprivation status of the students over the entire distribution rather than just the most deprived families. The coefficients on these indicators weaken as the deprivation decile increases from the most deprived students (decile 1) to the least deprived[5], but remain significant in all cases. The implication is that the more deprived a student's background, the more likely they are to opt for a vocational trajectory over an academic one.

Comparing the estimated effect of peer ability with the effects of other factors in our model suggests that peers have a substantial impact on education pathways at age 16. The strong link between individual ability and educational choice is well known but our results suggest the impact of peers ability is almost half as big as the impact of ones own ability (0.0018 compared to 0.0034). Similarly, the link between deprivation and educational outcomes is well known but our results show that a one percentile increase in the ability of ones peers has about one tenth of the impact of moving from the 10^{th} to the 7^{th} decile of the deprivation distribution, as measured by the IMD (0.00182 compared to 0.0176 in specification 2 of table 4).

[4]Rather than express the impact of a one percentile increase in peer ability, Mendolia et al. (2016) states the impact in terms of a 1 standard deviation increase. From our model specification which looks at vocational and A-Level results only, because the standard deviation of the percentile variable is 28.7, the impact of 1 standard deviation increase in peer ability is associated with a 6 percentage point decrease in the likelihood of selecting a vocational programme (or a 6 percentage point increase in the likelihood of a selecting an A-Level programme). This is broadly consistent with the evidence in Mendolia et al. (2016) which finds a one standard deviation increase in peer ability (measured as the percentage of peers who failed to achieve basic maths at KS3) is associated with an 8 percentage points increase in the likelihood of taking A-Levels.

[5]Decile 10 (least deprived) was omitted as the reference category

Table 4: GCSE achievers with Local Authority fixed effects (2011 cohort)

	All routes		Vocational vs academic only	
	(1) OLS	(2) IV	(3) OLS	(4) IV
KS2 test scores	-0.00350***	-0.00342***	-0.00433***	-0.00424***
	(0.0000605)	(0.0000657)	(0.0000678)	(0.0000736)
Peers KS2 test scores	-0.000648***	-0.00182***	-0.000760***	-0.00201***
	(0.0000732)	(0.000320)	(0.0000817)	(0.000350)
Female	-0.0374***	-0.0381***	-0.0467***	-0.0475***
	(0.00195)	(0.00202)	(0.00220)	(0.00228)
Ethnicity - Black	-0.108***	-0.112***	-0.131***	-0.135***
	(0.00492)	(0.00519)	(0.00561)	(0.00585)
Ethnicity - Asian	-0.163***	-0.173***	-0.188***	-0.199***
	(0.00429)	(0.00516)	(0.00478)	(0.00569)
Ethnicity - Other	-0.0491***	-0.0504***	-0.0599***	-0.0611***
	(0.00378)	(0.00408)	(0.00421)	(0.00453)
Mainstream school	-0.110	-0.0990	-0.130	-0.117
	(0.235)	(0.244)	(0.260)	(0.270)
Entered all sciences	-0.0764***	-0.0765***	-0.0937***	-0.0937***
	(0.00225)	(0.00232)	(0.00256)	(0.00262)
Entered at least one language	-0.0743***	-0.0641***	-0.0921***	-0.0811***
	(0.00256)	(0.00338)	(0.00294)	(0.00382)
Absences (% of sessions)	0.645***	0.691***	0.915***	0.970***
	(0.0210)	(0.0234)	(0.0265)	(0.0292)
IMD Decile 1	0.0670***	0.0328***	0.0809***	0.0450***
	(0.00526)	(0.0107)	(0.00584)	(0.0117)
IMD Decile 2	0.0640***	0.0350***	0.0767***	0.0461***
	(0.00456)	(0.00892)	(0.00503)	(0.00981)
IMD Decile 3	0.0564***	0.0321***	0.0679***	0.0422***
	(0.00434)	(0.00782)	(0.00484)	(0.00871)
IMD Decile 4	0.0465***	0.0256***	0.0573***	0.0352***
	(0.00419)	(0.00708)	(0.00467)	(0.00785)
IMD Decile 5	0.0379***	0.0205***	0.0471***	0.0287***
	(0.00395)	(0.00604)	(0.00439)	(0.00665)
IMD Decile 6	0.0367***	0.0236***	0.0445***	0.0307***
	(0.00385)	(0.00537)	(0.00428)	(0.00585)
IMD Decile 7	0.0279***	0.0176***	0.0339***	0.0227***
	(0.00362)	(0.00470)	(0.00407)	(0.00522)
IMD Decile 8	0.0226***	0.0153***	0.0275***	0.0200***
	(0.00353)	(0.00430)	(0.00395)	(0.00479)
IMD Decile 9	0.0174***	0.0122***	0.0210***	0.0155***
	(0.00328)	(0.00368)	(0.00368)	(0.00412)
Constant	0.000564	0.000762	0.0367***	0.0369***
	(0.00171)	(0.00179)	(0.00195)	(0.00203)
LA Fixed Effects	Yes	Yes	Yes	Yes
Observations	222,692	217,500	192,033	187,543

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

In order to test the validity of the assumptions required for an IV, a falsification test is implemented in Table 6. This test involves the allocation of students to artificial peer groups. If the exclusion restriction is valid, and it is the peers themselves rather than outside factors that are driving the decision to invest in a vocational qualification, then the coefficient on the Peers KS2 test scores variable should become insignificant. This is exactly what occurs, whilst the other coefficients remain unchanged.

4 Conclusion

Using the full census of English data for secondary school leavers at the age of 16 (following their GCSEs), this paper investigates the factors driving the decision to study for academic or vocational qualifications. More specifically, we aimed to find out whether the decision of young people to decide on further studies after the end of secondary schooling is influenced by their peers. Peer effects refer to the influence of a group on an individual and arise from the observation that group members are inclined to behave in a similar manner. In our administrative data, we identified peer groups as year groups in secondary schools because teaching is organised by subject groups, making it likely that the pupils will encounter the majority of the other students in their cohort at some point in their secondary school education.

In studies focusing on the influence of peers on individual decisions, one needs to address various sources of endogeneity. Firstly, individuals typically endogenously sort into groups, so that people, who show similar behaviour form groups, which – if unaddressed – might overstate the peer influence. Secondly, all members of the group influence each other at the same time, so it is not clear whether the actions and attributes of an individual's peers influence them or imitate them. In order to overcome such problems of endogeneity, a source of exogenous variation or a random allocation of peer groups needs to be found to estimate an unbiased effect of the peer group influence.

In this paper, we take advantage of the variation in peer groups that arises when students move from primary to secondary school to estimate the impact of secondary school peers on the choice of educational trajectory by age 16 (i.e. academic, vocational, a combination of the two, or no education at all). This analysis is limited to pupils with at least five A*-C GCSE, who have the choice to go on to Sixth Form or Further Education Colleges (and other Level 3 vocational education). A factor which might influences the ability of their peers but not their own ability is calculated using the KS2 test scores of people who went to the same primary schools as their peers but a different secondary school. These 'peers of peers' did not go to the same primary or secondary school as the individual of interest and cannot have had any direct impact on them, but did influence their peers during the period of primary schooling.

In order to test whether the identification strategy is robust, we created random secondary schools with random peers of peers in primary education and estimated the same equation. As expected, this falsification test showed no significant influence of the peers on the specific post-16 choice.

Our main results indicate that the composition of secondary school peers is an important determinant of a student's post-16 education choice. We find that while the main driver of educational choice is individual ability (expressed by the individual KS2 scores) and young people with higher ability more often decide in favour of academic rather than vocational education, peers have a significant impact too. The more able one's peers

are, the less likely one is to choose a vocational course after completion of their GCSEs, after controlling for the individual's own ability. In the models, we further control for absence in the final year of KS4, subjects entered at GCSE, ethnic group, gender and local authority fixed effects, all of which have a significant influence on choosing vocational instead of academic routes.

Another important finding in this paper is the influence of the social gradient on educational decision making. Instead of using Free School Meal (FSM) eligibility as an indicator for young people with a disadvantaged family background, we included low-level geographical deprivation data (Index of Multiple Deprivation, IMD). While this is not an individual-level covariate, the geographical areas represented are small and likely to represent household deprivation well on average. More importantly, the deciles of the IMD represent the full range of well off to the most deprived households and therefore offer a wider picture of deprivation than the FSM variable. For this variable, we see a significant relationship between deprivation and education choice over the whole distribution of wealth/IMD, i.e. students from more deprived families are significantly more likely to choose a vocational course, other things being equal, than those from better off families.

We have two main policy conclusions. Firstly, we evidence significant influence of the performance of peers on post-16 education choices for young people, e.g. people choose vocational education while higher ability would make then potentially more successful in a Sixth Form College (or the other way round). Ideally, the influence of peers ability should be zero to allow people to 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 and better information on educational choices needs to be offered at age 16 so that individuals can make impartial choices.

Secondly, our findings also tie in with the current policy aim of supporting the "just-about-managing (JAM) families". The results of this paper indicate that students from such families are less likely to enrol in an academic course irrespective of their ability and the ability of their peers, potentially limiting their future options.

References

Ashworth, J. & Evans, J. (2001), 'Modelling student subject choice at secondary and tertiary level: A cross-section study', *The Journal of Economic Education* **32**(4), 311–320.

Bayer, P., Hjalmarsson, R. & Pozen, D. (2009), 'Building criminal capital behind bars: Peer effects in juvenile corrections', *The Quarterly Journal of Economics* **124**(1), 105–147.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**(1), 41–55.

Burke, M. & Sass, T. (2013), 'Classroom peer effects and student achievement', *Journal of Labor Economics* **31**(1), 51–82.

Calvó-Armengol, A., Patacchini, E. & Zenou, Y. (2009), 'Peer effects and social networks in education', *The Review of Economic Studies* **76**, 1239–1267.

Carrell, S., Fullerton, R. & West, J. (2009), 'Does your cohort matter? Measuring peer effects in college achievement', *Journal of Labor Economics* **27**(3), 439–464.

Cipollone, P. & Rosolia, A. (2007), 'Social interactions in high school: Lessons from an earthquake', *American Economic Review* **97**(3), 948–965.

De Giorgi, G., Pellizzari, M. & Redaelli, S. (2010), 'Identification of social interactions through partially overlapping peer groups', *American Economic Journal: Applied Economics* **2**(2), 241–275.

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

Gibbons, S. & Telhaj, S. (2015), 'Peer effects: Evidence from secondary school transition in england', *Oxford Bulletin of Economics and Statistics* **59**(1), 548–575.

Graham, B. (2008), 'Identifying social interactions through conditional variance restrictions', *Econometrica* **76**(3), 643–660.

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

Kiss, D. (2013), 'The impact of peer achievement and peer heterogeneity on own achievement growth: Evidence from school transitions', *Economics of Education Review* **37**, 58–65.

Lahno, A. & Serra-Garcia, M. (2015), 'Peer effects in risk taking: Envy or conformity?', *Journal of Risk and Uncertainty* **50**, 73–95.

Lavy, V., Silva, O. & Weinhardt, F. (2012), 'The good, the bad, and the average: Evidence on ability peer effects in schools', *Journal of Labor Economics* **30**(2), 367–414.

Manski, C. (1993), 'Identification of endogenous social effects: The reflection problem', *Review of Economic Studies* **60**(3), 531–542.

Mas, A. & Moretti, E. (2009), 'Peers at work', *American Economic Review* **99**(1), 112–145.

Mendolia, S., Paloyo, A. & Walker, I. (2016), 'Heterogeneous effects of high school peers on educational outcomes', *IZA Discussion Paper* pp. 1–38.

Poldin, O., Valeeva, D. & Yudkevich, M. (2015), 'Choice of specialization: Do peers matter?', *Applied Economics* **47**(44), 4728–4740.

Sacerdote, B. (2001), 'Peer effects with random assignment: Results for dartmouth roommates', *Quarterly Journal of Economics* **116**(2), 681–704.

Stinebrickner, R. & Stinebrickner, T. (2006), '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**, 1435–1454.

Trogdon, J., Nonnemaker, J. & Pais, J. (2008), 'Peer effects in adolescent overweight', *Journal of Health Economics* **27**, 1388–1399.

Vardardottir, A. (2013), 'Peer effects and academic achievement: A regression discontinuity approach', *Economics of Education Review* **36**, 108–121.

Zimmerman, D. (2003), 'Peer effects in academic outcomes: Evidence from a natural experiment', *Review of Economics and Statistics* **85**(1), 9–23.

A Data Set Up

A.1 National Pupil Database

The National Pupil Database (NPD) provides information about students and their academic achievements in English schools, which covers Key Stage 1 to Key Stage 5. The analysis in this paper uses NPD data for three consecutive cohorts; those completing Key Stage 4 in the academic years 2010/2011, 2011/2012 and 2012/2013. This dataset is used to identify peer groups at primary and secondary schools (and consequently the transition groups), as well as prior ability (measured prior to the transition at the end of Key Stage 2) and GCSE achievement. Additionally, the Key Stage 5 information identifies the choice for students who choose to undertake A levels after the completion of compulsory schooling.

A.1.1 Key Stages 1-4

In order to deal with selection problems, only schools which were specifically stated to be comprehensive were included. The NPD includes flags for those who joined in year 9 and those who joined in year 7 or 8 outside of July-September. These were dropped to prevent sorting. Additionally, only individuals who completed year 11 at the same time were included as the decision is made at this stage. Finally, only individuals who completed the other Key Stages at the expected time were retained to ensure that the peer group measure is accurate.

Duplicates were then dealt with in the following order[6]:

1. Where one observation is counted in the LA results calculation and the duplicate isn't, the duplicate is dropped.
2. Where one observation is counted in the school results calculation and the duplicate isn't, the duplicate is dropped.
3. Where one observation is counted in the national results calculation and the duplicate isn't, the duplicate is dropped.
4. Priority is given to observations which are counted in the 'number on roll' calculation, then those 'ending compulsory schooling here' over observations which aren't.
5. The observation with the highest number of GCSE entries is kept as this is where they are most likely to have completed compulsory schooling.
6. The observation with the highest number of full GCSE entries is kept.
7. The observation with the highest number of short GCSE entries is kept.

Finally, only secondary school cohorts of at least ten students were kept in order to create a realistic peer group measure.

[6] At this stage all duplicates originate from KS4 so the observation where they completed KS4 was retained as this gives GCSE achievement. This does not have implications for identification as peer groups are measured at KS2 and this information is not used other than for GCSE achievement. Additionally, GCSE achievement is preserved for dropped observations through the calculation of a total measure.

A.1.2 Key Stage 5

In order to identify the initial choice, only the earliest academic year for each learner was kept[7]. Additionally, given that the analysis is concerned with the choice of students who have just completed year 11, only observations for year 12 students were kept.

Duplicates arose for individuals who changed schools during the academic year. The best method identified to deal with this was to retain the observation where the provider identifier is equal to that of the previous year (it was considered more likely that the individual transferred away from the previous institution than back towards it).

A.2 Individualised Learner Record

The Individualised Learner Record (ILR) is an administrative dataset which contains information on learner participation and achievement in the Further Education sector. This dataset is used to identify the choice for individuals who choose a more vocational route, as well as for those who take A levels via a further education provider. Given that the analysis uses NPD cohorts which complete compulsory schooling in 2010/2011, 2011/2012 and 2012/2013, their choice will be realised in the following academic year. Hence ILR cohorts for the academic years 2011/2012, 2012/2013 and 2013/2014 are used. The ILR was merged with the Learning Aims Database (LAD) in order to gain more information about the courses and learning aims that the learners were enrolled in.

The ILR aims file provides information about every aim that an individual undertakes, rather than the course that they are enrolled in, so there can be up to 80 observations per person. Given that the analysis is only interested in the initial choice, only observations which began in the earliest academic year were retained[8]. Furthermore, only observations which began straight after KS4 (August, September or October) were kept to minimise the time available for factors other than secondary school peer group to intervene in the choice. Courses which began after this time period will be equated with making no choice at all so, for the purposes of this analysis, no choice should be taken to mean no immediate choice.

Duplicates were subsequently dealt with in the following way:

1. Where an individual was enrolled in aims at more than one institution, the start and end dates (both planned and actual) of the aims were used to identify any drop outs. If an aim at one provider began after the end date of the previous aim at another provider, then this observation was dropped[9]. The same method was subsequently employed for the cases where individuals appear to transfer aims within the same institution.

2. Aims which are purely supplementary were dropped as these cannot be considered a genuine choice. For example, 'Tutorial and enrichment studies for full time 16-18 students studying for example, GCE A levels; AS levels; GCSEs and short course

[7]Given that the majority of courses that appear in the Key Stage 5 data last for around two years, many students appear twice (e.g. once for AS levels and again for A2 levels).

[8]When this is merged with the NPD, the academic year is used to ensure that individuals began in the immediate academic year as opposed to taking a gap year etc.

[9]This was necessary as opposed to keeping the institution with the earliest start date to prevent the loss of aims in the case where an individual is genuinely enrolled in aims at several institutions at the same time.

GCSEs' is a course which is simply an accessory to A levels or GCSEs. Similarly, 'Preparation for Life and Work' was dropped if there was a better alternative.

3. Aims which did not lead to a recognised qualification or are not externally certified were also dropped as these were also deemed to not be a genuine choice. Examples of the former include 'Foundation Learning Weekly Learning Aim' and 'Participant receiving IAG', the latter includes 'basic literacy course' and similar aims.

4. For aims which began at the same time, the aim with the highest number of guided learning hours was kept as this was likely to be the main choice[10].

5. Observations of less than two weeks planned duration were dropped if other aims are longer as this is likely to simply be a pre-requisite to the main aim or an enrichment activity. It was considered unlikely that people start college with the sole intention of a two week course if they start another longer aim around the same time.

6. For individuals who repeat the same qualification or are asserted to have the same aim twice, the initial instance according to start date was retained.

7. 'Functional skills' aims were dropped if there were alternatives as the main purpose of these courses is simply to improve Maths, English and ICT skills if prior attainment is low.

8. In the case that an individual had more than four aims after the steps undertaken above, any subsequent aims were dropped if the start date was after that of the initial aims.

9. Where learners are reported to be taking both AS levels and A2 levels, AS levels were kept (in the case that there were more than four other aims) as these must be taken first.

After these steps, learners still retained up to eight aims. In order not to lose any information, the dataset was then reshaped to allow for the inclusion of the remaining aims in the final dataset.

[10]This measure does not necessarily prevent learners from having several main aims. For example, an individual taking four A levels would retain this information as the number of guided learning hours would be equal.

A.3 Qualification Classification

Post-compulsory qualifications were classified through a mixture of the aim type category and a string search of the aim title. It was necessary to use both of these aspects to avoid any misclassification: for example, the category 'GCE A level' also includes Applied A levels which are considered as a vocational alternative and would have been misspecified as academic if the aim title was not also taken into account. Qualifications were classified in the following way:

Table 5: Classification of main post-compulsory qualifications

Academic	**Vocational**	**Basic Skills**	**Other**
A Levels	BTECs	Key Skills	Extended Project
IB Diploma	NVQs	Functional Skills	FE Enrichment
Access to HE Diploma	Applied A Levels	GCSEs	First Aid
Pre-U Certificate	14-19 Diploma	Basic Literacy	Higher project
etc.	etc.	etc.	etc.

Note: This is not an exhaustive list, the remaining courses were classified according to best judgement.

B Placebo-tests

Table 6: Falsification Test - GCSE achievers with Local Authority fixed effects (2011 cohort)

	All routes		Vocational vs academic only	
	(1) OLS	(2) IV	(3) OLS	(4) IV
KS2 test scores	-0.00354***	-0.00357***	-0.00439***	-0.00441***
	(0.0000401)	(0.0000457)	(0.0000449)	(0.0000525)
Peers KS2 test scores	0.0000313	-0.00423	0.0000425	-0.00476
	(0.0000270)	(0.00282)	(0.0000298)	(0.00356)
Female	-0.0370***	-0.0366***	-0.0461***	-0.0454***
	(0.00158)	(0.00167)	(0.00176)	(0.00193)
Ethnicity - Black	-0.106***	-0.108***	-0.128***	-0.130***
	(0.00389)	(0.00455)	(0.00441)	(0.00495)
Ethnicity - Asian	-0.157***	-0.156***	-0.181***	-0.179***
	(0.00270)	(0.00310)	(0.00307)	(0.00385)
Ethnicity - Other	-0.0485***	-0.0474***	-0.0594***	-0.0576***
	(0.00335)	(0.00367)	(0.00379)	(0.00432)
Mainstream school	-0.117	-0.0617	-0.136	-0.0895
	(0.237)	(0.240)	(0.261)	(0.268)
Entered all sciences	-0.0763***	-0.0762***	-0.0937***	-0.0937***
	(0.00151)	(0.00162)	(0.00168)	(0.00182)
Entered at least one language	-0.0792***	-0.0782***	-0.0978***	-0.0967***
	(0.00171)	(0.00195)	(0.00192)	(0.00226)
Absences (% of sessions)	0.623***	0.627***	0.887***	0.890***
	(0.0176)	(0.0187)	(0.0209)	(0.0222)
IMD Decile 1	0.0861***	0.0867***	0.103***	0.104***
	(0.00414)	(0.00434)	(0.00465)	(0.00489)
IMD Decile 2	0.0800***	0.0804***	0.0955***	0.0951***
	(0.00378)	(0.00400)	(0.00420)	(0.00451)
IMD Decile 3	0.0702***	0.0700***	0.0839***	0.0831***
	(0.00358)	(0.00377)	(0.00403)	(0.00435)
IMD Decile 4	0.0581***	0.0586***	0.0709***	0.0712***
	(0.00345)	(0.00364)	(0.00380)	(0.00406)
IMD Decile 5	0.0472***	0.0459***	0.0578***	0.0562***
	(0.00329)	(0.00360)	(0.00366)	(0.00412)
IMD Decile 6	0.0442***	0.0446***	0.0532***	0.0538***
	(0.00312)	(0.00334)	(0.00348)	(0.00380)
IMD Decile 7	0.0337***	0.0341***	0.0406***	0.0409***
	(0.00294)	(0.00312)	(0.00330)	(0.00352)
IMD Decile 8	0.0270***	0.0291***	0.0327***	0.0350***
	(0.00297)	(0.00344)	(0.00328)	(0.00390)
IMD Decile 9	0.0204***	0.0213***	0.0246***	0.0251***
	(0.00277)	(0.00301)	(0.00306)	(0.00333)
Constant	0.000511	0.000434	0.0366***	0.0365***
	(0.000922)	(0.00238)	(0.00104)	(0.00270)
LA Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.1088	0.0049	0.1501	0.0317
Observations	222,627	222,623	191,978	191,974

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