Mismatch in higher education: prevalence, drivers and outcomes

Stuart Campbell, Lindsey Macmillan and Gill Wyness

December 2019

UCL Institute of Education

Nuffield Foundation

Acknowledgements

This report summarises research from a study funded by the Nuffield Foundation (grant reference EDU/172585) entitled *Undermatch in higher education: prevalence, drivers and outcomes*. The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org

Nuffield Foundation

We thank Ellen Austin, Vikki Boliver Paul Clark, Paul Gregg, Tim Leunig, Sandra McNally, and Donna Ward for helpful comments, as well as seminar participants at Columbia University, LSE's Centre for Economic Performance, ISER, and Queen's Belfast, workshop participants at Catanzaro, IAB, and York, and conference participants at APPAM, EALE, ESPE, RES, and SOLE. We also thank the Nuffield foundation, especially Cheryl Lloyd. We use data from NPD-HESA-DLHE provided by DfE, (ref DR170721.02), and data from Next Steps provided by UCL (2018).The views expressed in this report are those of the authors, and all errors and omissions remain the sole responsibility of the authors.

Contents

1 Executive summary

Overview

We consider the extent of mismatch, students attending courses that are less or more selective than might be expected given their academic attainment, in UK higher education (HE) for the first time. There is significant under- and over-match in the UK, with 15% (23%) of students under- and over-matching when we measure course quality[1] based on course-level attainment (future graduate earnings). There are substantial socio-economic status (SES) and gender gaps in mismatch, with low SES students and women attending lower quality courses than their attainment might otherwise suggest. This has important implications for social mobility and the gender pay gap. While subject choice at university is a key driver of the gender gap, the SES gap can largely be explained by the secondary school attended. There is also an interesting geographical dimension with low SES students who travel to attend university increasingly likely to match in the same way that high SES students do, but for students who stay close to home, the SES gaps are striking. There are penalties to undermatching in terms of degree outcomes, and in the labour market, with those students who undermatch less likely to achieve a first or a 2:1, and likely to earn less 3.5 years after graduation. Conversely, those who overmatch achieve more positive outcomes than similar matched students. This research suggests that there is an important role for information, advice and guidance, and university outreach programmes, to ensure that students are making informed choices.

1.1 Introduction

Increasing enrolment in HE is a preoccupation of governments around the world. As a result, much academic research has been devoted to examining policies intended to increase university participation, particularly to under-represented groups such as those from low SES backgrounds. However, less attention has been given to the types of universities and courses students enrol in once they decide to continue with their education, and in particular the 'match' between the student and their course.

Recent research from the US (Dillon and Smith, 2017, 2019; Black et al, 2015, Smith et al, 2013) has begun to focus on the phenomenon of 'mismatch'. Although there is some variation in the exact definitions, mismatch broadly consists of 'undermatch', where students attend

[1] We define course quality in two ways, as will be discussed below. These definitions are of course subjective, but are useful for ranking courses.

universities that are less selective than might be expected, given their academic credentials, and 'overmatch', where students attend universities that are more selective than might be expected, given their credentials.

Existing evidence (which is, to date, exclusively from the US) suggests that a significant proportion of students are undermatched, and that undermatch is more common among ethnic minority students and those from low socio-economic (SES) backgrounds (Smith et al, 2013; Dillon and Smith, 2017; Dillon and Smith, 2019). This research also suggests that these students are less likely to overmatch than their more advantaged counterparts, even when similarly qualified. Despite the growing research base in the US, there is a paucity of research on mismatch in the UK context.

Given the well-documented returns to high status universities and subjects (Belfield et al, 2018), understanding the extent to which disadvantaged students are less likely to enrol in high quality courses than more advantaged students of similar academic attainment, is important for equity and social mobility.

Issues regarding the university admissions process and its impact on disadvantaged students are at the forefront of the current HE policy debate. Universities UK has launched a review of the admissions process, whilst the Office for Students (OfS) is to launch a review soon. The Labour Party have announced plans to radically reform HE admissions by scrapping university offers based on predicted grades and implementing a new system of post-qualification admissions (PQA), aiming to make the admissions process fairer. The Conservative Party have issued a statement backing the OfS review, and endorsing the opportunity to consider the pros and cons of potential models of PQA. Our research will help policymakers understand the extent to which the current system is fair in terms of the courses that students from different groups attend – with potential implications for information, advice, and guidance during the admissions process – and hence is highly relevant to this debate.

1.2 Project Aims

First, we aim to understand the extent of mismatch in the UK HE system. We create two measures of student-to-university match. Our first measure matches students to courses based on their academic attainment, and the academic attainment of their fellow students, measuring the extent to which students attend courses that are commensurate with their academic attainment up to that point. Our second measure matches students to courses based on the average earnings of graduates from that course, measuring the extent to which students attend courses that are likely to generate future earnings commensurate with their position on the academic attainment distribution.

Our second aim is to document the characteristics of mismatched students. We might have reason to believe that low SES students may be more likely to undermatch than their counterparts from richer backgrounds. For example, such students may have less information about the benefits of attending a more academically prestigious institution or an institution associated with higher earnings. Alternatively, they may have less access to social and cultural capital (Britton et al, 2016).

Our third and fourth aims are to evaluate the impact of mismatch on university and labour market outcomes. Overmatch may at first glance appear to be beneficial to the student as they will be attending a higher quality course than expected. However, overmatched students may struggle to keep up with their better prepared peers and the material being taught, potentially resulting in lower graduation rates. Similarly, being undermatched could be equally harmful, as students will be attending HE courses with fewer financial resources and lower prestige, potentially impacting graduation rates and future earnings. On the other hand, such students may benefit from being "big fish in a small pond." (Murphy and Weinhardt, 2018). Thus understanding the consequences of mismatch is an important empirical question.

1.3 Previous Literature

The small body of literature in this area comes exclusively from the US. These papers find a high degree of student to university mismatch, with estimates suggesting that around 25% of students are mismatched in the US (Dillon & Smith, 2017).

Several studies have revealed that disadvantaged students are more likely to undermatch and less likely to overmatch (Hoxby and Avery, 2012; Smith et al, 2013; Dillon & Smith, 2017), and that information, geographical isolation and financial constraints are important drivers of match quality.

A small number of studies have also examined the consequences of mismatch for future outcomes (Arcidiacono & Lovenheim, 2016; Dillon & Smith, 2017), finding some evidence that the fit between the student and the university matters. For example, Dillon and Smith (2017) show that more able students benefit more from being at a matched (i.e. high quality) university in terms of time to degree. There is also some evidence that the university fit matters for later earnings outcomes. In a causal study of the impacts of overmatch, Arcidiacono et al (2014) focus on the affirmative-action ban in California (Proposition 209). They find that overall graduation rates for minorities in the University of California (UC) system increased by over 4 percentage points after the ban. After the ban, minority students – who were less academically prepared – were "reshuffled" towards less selective universities. Their positive finding implies that these less selective institutions are better at graduating less well-prepared students. However there is a paucity of research on gender gaps in mismatch, and on mismatch in a non-US context.

1.4 Defining and Measuring Mismatch

Our focus is on the cohort of young people who took their General Certificate of Secondary Education (GCSE) exams in 2006 and their Advanced level qualifications (A level) or equivalent exams two years later in 2008, at the end of secondary school, before entering HE either straight away or after one gap year.

- We use individual-level linked administrative data from schools and universities, and aggregate data on graduate earnings from tax records to construct our measures of match, to understand the characteristics of mismatched students, and to evaluate their degree outcomes.
- We use linked survey data on these students' labour market outcomes to understand the consequences of match for employment and earnings.

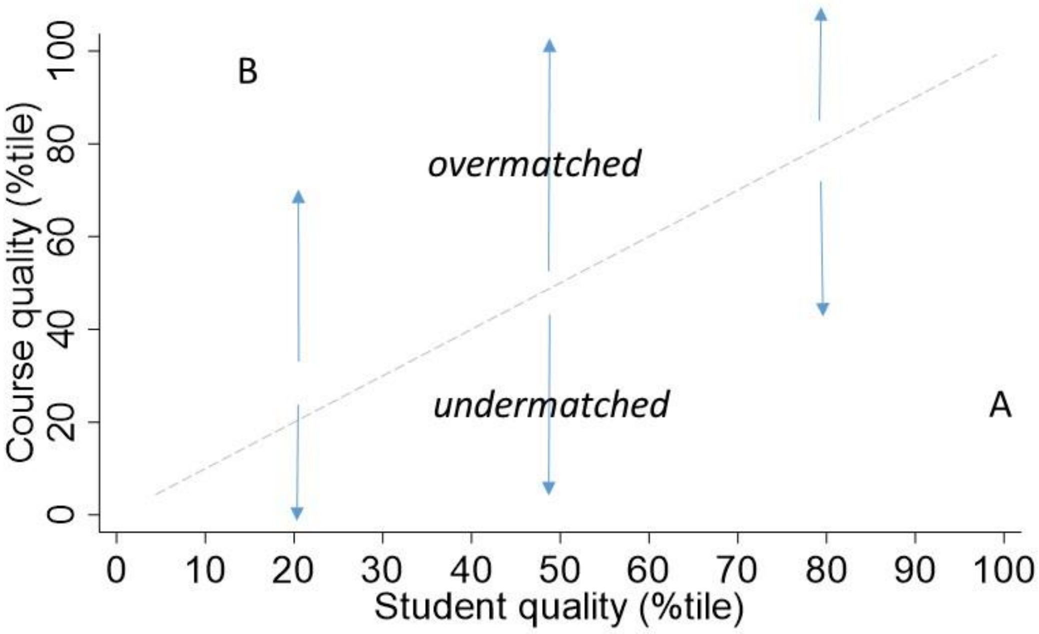
Figure 1 illustrates how we construct our measures of match.

- We calculate each students' percentile in the national academic attainment distribution (based on their best 3 A levels or equivalents, weighted for subject difficulty[2]). A student at the top of the distribution (e.g. someone with 3 As in Maths, Advanced Maths, and Physics) will be at the 100th percentile on the x-axis of Figure 1. A student at the bottom of the distribution will be at 0 on the x-axis[3].
- We then calculate each course's position on the national course quality distribution (of all university courses – some 1,300), based on either the grades of students taking that course (points-based match) or the earnings of a previous cohort of graduates 5 years

[2] For more details on subject difficulty weighting, see Section 5.
[3] Note that our data includes every pupil in the state school population, but excludes those at independent schools. Thus each pupil will be ranked in comparison to other state school pupils only.

after graduation (earnings match). Here, low quality courses (courses populated with low-attaining students, or with low associated earnings) are at 0 on the y-axis, and high quality courses (high-attaining students/high graduate earnings) are at 100 on the y-axis.

Figure 1: Measuring mismatch



- Mismatch is calculated by comparing each student's position with that of their chosen course, subtracting the individual's percentile from the course percentile. When a student attends a course that is at roughly the same percentile on the course quality distribution as his/her percentile on the attainment distribution, they will be on the 45 degree dotted line (match index of approx. zero).
- Students are defined as undermatched if they attend a course that is at a lower percentile on the course quality distribution than their own percentile on the student quality distribution (a negative match index). In other words, where the student has higher academic attainment than their fellow students on the same course.
- Students are defined as overmatched, meanwhile, if they attend a course that is at a higher percentile on the quality distribution than their own percentile (a positive match index). In other words, where the student is lower attaining than their course peers. While we prefer to measure mismatch as an index, in order to be comparable with the

US, at some points in this report we adopt the binary definition of mismatch from Dillon and Smith, 2017 where mismatch of is +/- 20 percentiles from the matched course.

We also consider the extremes of under and overmatch. For example, a student at point A in Figure 1 is extremely undermatched. He/she is at the top of the student attainment distribution, but is attending a course at around the 20th percentile – a match index of -80. The student at point B is severely overmatched, being at the 20th percentile in the attainment distribution, but is attending a course at around the 90th percentile – a match index of +70. While these extremes are of course rare in reality, we illustrate the full distribution of match in Figure 2, and focus on extremes of match in Section 7.

1.5 Key Findings

We find substantial amounts of both undermatch and overmatch. 15% of students are overmatched and 15% are undermatched using our points-based measure[4]. For our earnings-based measure 23% overmatch and 23% undermatch. Dillon and Smith (2017) find around 25% of students in the US are overmatched and 25% undermatched according to their composite college-input-quality measure. This is most comparable to our points-based measure of match, and while it would problematic to draw strong conclusions from this (as will be discussed in Section 6), the comparison is suggestive that there is more mismatch in the US than in the UK.

There are large social and gender gradients in mismatch.

- We find that students from low SES backgrounds are more likely to undermatch, and less likely to overmatch than their high SES counterparts. This is true for both our measures of match – meaning that low SES students attend courses that are less academically selective, and with lower average earning five years after graduation, than their high SES counterparts, even when they have similar levels of prior attainment. This has important implications for social mobility.
- By contrast, we find that while women and men attend equally academically selective courses, women enrol in courses with substantially lower average earnings than men, even when they have similar prior attainment. This is potentially important for understanding the gender pay gap.

[4] Where students are defined as undermatched if they are ranked 20 percentiles below their course, and undermatched if they are ranked 20 percentiles above.

- While gender gaps in mismatch are largely explained by subject studied at university, SES gaps are driven by secondary school attended. In addition, low SES students who travel to attend university look increasingly similar in terms of match to high SES students, but large SES gaps remain between students who study at universities close to home.

We also find that White students are more likely to undermatch relative to ethnic minority students, and that students who prefer to live closer to home, who are less certain about whether they will go to university or not by age 16, and students who do not get into university with their first choice of subject, are all more likely to undermatch.

Moreover, we find that mismatch matters for later outcomes. There are economic penalties associated with undermatching for both university performance and labour market outcomes; students undermatched on points are more likely to drop out of university, and to get a lower class degree (less than a 2.1)[5], and go on to earn less in the labour market than similarly qualified but matched students. Those who overmatch on points, meanwhile, are less likely to get a lower class degree, and go on to earn more in the labour market.

These economic penalties for undermatched students, coupled with the social and gender gradients in mismatch, imply that undermatch impedes social mobility and gender equality, and has implications for the gender pay gap.

As described, our results refer to the group of students who entered university in 2008. Since then, the proportion of students going to university has increased, and there has been increased activity devoted to widening participation. However, the gap between rich and poor students has remained relatively stable since 2008, suggesting our findings are reasonably representative of the current situation.

1.6 Who are the mismatched?

We find students from lower socio-economic groups (where we measure SES using a composite measure based on free school meals status, plus area-based measures of deprivation) are more likely to undermatch, and less likely to overmatch throughout the attainment distribution.

[5] We choose this particular outcome as it is widely accepted in the UK labour market that achievement at the level of 2.1 or above is a key differentiator for employers. Indeed, graduates with a first or 2.1 have been shown to earn around 8% more than those with lower class degrees (Feng & Graetz, 2015; Naylor et al 2015; Walker & Zhu, 2013). This is also often the minimum requirement for entry to graduate programmes.

In addition we highlight a significant gender gap in match. While women and men attend equally academically selective courses, women enrol in courses with substantially lower expected earnings, conditional on prior attainment. These socio-economic and gender inequalities become even starker among the extremely undermatched. High-attaining White, Black Carribean and Black students from other backgrounds also undermatch to a greater extent than students from all other ethnic minorities.

These gaps are particularly pronounced for our earnings-based measure of match, where ethnic minority students overmatch more and undermatch less (again with the exception of Black Caribbean and Black students from other backgrounds) than White students.

1.7 Drivers of mismatch

Accounting for the subject of degree does not reduce the socio-economic gap in match. In other words, when students are of similar attainment and studying the same subject, low SES students study at lower quality institutions. Thus, we can conclude that a key driver of SES inequalities in match is the institution attended.

However, subject choice does account for most of the gender gap in earnings match; the fact that women attend courses with lower earnings potential than men is largely driven by the subjects that women choose, rather than the institution. But a gap remains for high-attaining women, implying that regardless of subject of study, these women attend universities with lower associated earnings. This implies that it may be important to provide women in particular with information about the economic returns associated with different subjects.

While we do not find any evidence that geography plays a role in driving the SES gap in match, we find interesting differences by distance travelled to university.

- High-attaining, low SES students are more likely to attend universities close to home, but those who do so are worse matched than high SES students who attend universities close to home. High-attaining low SES students going to universities near home tend to choose a post-1992 institution, whereas high-attaining high SES students staying near home tend to choose a nearby Russell Group university.
- The fact that low SES students attend universities closer to home could be driven by information constraints, or fear of not fitting in at universities typically attended by more advantaged students.

- Interestingly, meanwhile, those low SES students who move away to attend university face no match penalty.

Further, there is a role for students' preferences, and forward planning in mismatch. Students who are more certain about whether they will go to university or not by age 16, and who have some idea of university prestige and its importance, are better matched suggesting that students who are on the path towards university, and who have undertaken some research, or have useful networks of people offering advice and guidance, are more likely to find a good match. Meanwhile, as described above, those who may be influenced by the location of a university (rather than its prestige) tend to be worse matched.

Students who do not get their first choice of subject are also more likely to undermatch in terms of the earnings measure of match. Students who reported (in survey data) that the subject they are currently studying at university was not their first choice, were found to be more likely to be undermatched on earnings. These students were rejected by their first choice of course, though we are unable to observe whether the first choice was a better match.

Note that we are unable to observe or take into account other elements of the university admissions process which may affect the process by which a student under or overmatches – such as where they apply, the use of their personal statements, or their performance at interviews. For example, students may be undermatched according to their grades, but a) may have not applied to a matched course, b) may have applied but had a very poor personal statement or poor interview performance, resulting in them ending up in a course below their apparent attainment level, c) applied and were offered a lower grade, or d) any other permutation of the admissions process.

Equally, a growing proportion of universities are allowing students from low SES backgrounds to enter with lower grades than required due to the use of contextual admissions (Boliver et al, 2017). Such students would be overmatched by our definition, but again we are unable to observe the extent to which contextual admissions may be important drivers here.

1.8 Outcomes of mismatch

Overmatched students typically have similar or more positive outcomes than matched students at university and beyond, while undermatched students typically have more negative outcomes than matched students. These differences are significant enough in magnitude to be economically important.

There is a penalty associated with being undermatched. After taking account of demographics, school factors and broad university group (Russell Group v non), students that are undermatched on our attainment measure are 4 percentage points more likely to get a lower-class degree than those that are well matched. While we cannot say why these negative effects occur, this supports the hypothesis that students who attend courses at less selective universities than they could have, pay the price in terms of having lower attaining peers and attending courses which have fewer resources. While being undermatched on our attainment measure is not associated with poorer labour market outcomes, being undermatched on earnings – essentially attending a course with lower future earnings than expected – does appear correlated with individual-level lower earnings (around 15% lower than a matched student) 3.5 years after graduation.

Conversely, there are positive associations with being overmatched. Students overmatched on our attainment measure are less likely to get a lower class degree, by around 3 percentage points, after conditioning on observable differences between matched and overmatched students. This suggests that students who enter courses with higher attaining peers do not appear to struggle academically. Being overmatched also appears to bear fruit in the labour market. Being overmatched, in terms of attainment and future earnings, is correlated with individual-level earnings around 6% higher than matched students 3.5 years after graduation.

1.9 Recommendations

Our findings have important implications for policy and society. Our finding that low SES students are attending courses with lower academic prestige, and lower associated earnings, regardless of prior attainment, has important implications for their future earnings. It is unlikely that credit constraints can explain these SES gaps – the vast majority of courses in the UK system charge the same fees, and students are all able to access loans for the full fee amount, and for support whilst they are at university. A more likely explanation is that low SES students have less (or lower quality) information available to them when making choices (or do not access information that is available to them). However, there are a number of processes at play in the UK's university application system, which could be responsible for the greater degree of undermatch among low SES students, such as the use of predicted grades in university applications.

Our finding that females attend courses with lower associated earnings than men on average has relevance for the much documented gender pay gap. Even among high attainers, we show that women are just as likely to enter academically competitive degree subjects as men, but appear to do so at less financially rewarding institutions. This suggests that research into the gender pay gap needs to focus on where women decide to study as well as what.

Our outcomes analysis offers suggestive evidence that undermatching has negative consequences in terms of university and labour market outcomes. Coupled with the characteristics of those who are more likely to undermatch, this has worrying implications for social mobility and highlights the importance of efforts to improve the quality of match of students to universities, rather than focusing on just getting students to attend.

Information, Advice and Guidance

The most obvious policy solution would be to improve the level and quality of information available to undermatched students. For example, students could be provided with information on the entry requirements and labour market returns to different courses at key decision-making ages (Belfield et al, 2018).

However, simply offering information (e.g. on the different returns associated with different institutions) may not be enough to resolve these issues. Studies have shown that those who gain the most from this type of information may be the least likely to consume it, and, to be effective, information has to be carefully targeted (McNally, 2016, Dynarski et al., 2018, Sanders et al., 2018, Sanders et al., 2017).

The UK Applications System

Given the characteristics of those who undermatch, it could be that our current applications system is creating some of the mismatch, as students apply to universities based on their predicted rather than actual grades. Wyness (2016) shows that high achieving low SES students are more likely to have predicted grades that *understate* their actual results.

Clearing may also play a role in mismatch. There is evidence that the number of students going through clearing is increasing, and also it is likely that students from different families may approach clearing differently (O'Kelly, 2019). The clearing system could reduce mismatch, since those who do go through clearing are applying to universities on the basis of their actual grades rather than their predicted grades. On the other hand, if capacity constraints mean that students within clearing have more limited options available, relative to waiting to apply the following year, then this could lead to more mismatch. Unfortunately our data does not allow us to observe whether or not students went through clearing so we are unable to isolate its role.

Low SES students are also typically more risk averse (Schurer, 2015), meaning that they are more likely to apply to courses that are easier to access in terms of grade requirements, rather than taking a risk on courses that may be harder to get in to. Both of these market failures (i.e. risk aversion and information failures) are potential routes for mismatch to occur.

A policy solution to minimise these issues would be to introduce PQA. As discussed, the Labour Party have announced that they would move to such a system should they gain power (Labour Party, 2019). Creating an admissions system based on observed rather than predicted grades at A level would eliminate the issue of under-predicting for low SES students and reduce risk aversion issues, as the decision would be based on real information. This would enable all students to match more effectively to courses.

Suggested intervention

Building on these policy solutions, we propose an intervention providing targeted information within the UCAS admissions system.

- Based on the idea of targeted advertising, we propose a new system which offers students course suggestions based on their A level (or equivalent) subjects and grades. This could either be the students' grade predictions, or preferably, if policy changes, their actual grades.
- This system could offer a range of filters such as degree subject preference (where students would pick their preferred subject, and would be offered suggested related matched courses), and location preference (where students would be offered suggested matched courses in the area of their choice).

This intervention would provide targeted IAG and, if coupled with PQA, could improve the quality of student to course match for those most at risk of mismatch.

A key benefit of this intervention is that, by working through the UCAS system, the vast majority of students would be reached. Whilst other information based interventions, such as current DfE projects (see https://www.gov.uk/government/news/winners-announced-for-new-student-apps) are offering innovative ways for students to access information during the decision-making process, these are likely to be opt-in only. As previous research has highlighted (McGuigan et al, 2016), the students most likely to use such as are those who are already well informed.

2 Introduction

Increasing enrolment in HE is a preoccupation of governments around the world. As a result, much academic research has been devoted to examining policies intended to increase participation by relaxing credit constraints (Murphy et. al. 2019), providing better information (McGuigan et. al. 2016) or improving prior academic attainment (Chowdry et. al. 2013). However, less attention has been given to the types of universities and courses students enrol in once they decide to continue with their education, and in particular, whether certain types of students are undermatching (attending courses that are less selective than they could, given their grades) or overmatching (attending courses that are more selective than they could, given their grades).

Evidence suggest that around 50% of students in the US are mismatched (Dillon and Smith, 2017) and that ethnic minority students and those from low socio-economic (SES) backgrounds are more likely to undermatch and less likely to overmatch (Hoxby and Avery, 2012; Smith et al, 2013; Dillon and Smith, 2017). However, no such evidence exists in the UK context.

The consequences of student to university mismatch are theoretically ambiguous. Overmatch may at first glance appear to be beneficial to the student, as they will be attending a higher quality college than expected. However, overmatched students may struggle to keep up with their better prepared peers and the material being taught, potentially resulting in lower graduation rates. Similarly, being undermatched could be equally harmful, as students will be attending less selective institutions. These institutions typically have smaller budgets and spending per student (on both academic and non-academic support), and lower numbers of staff to students. For example, student staff ratios are around 11 students per staff member at Oxbridge universities, versus around 15 or higher for low-ranked universities (The Guardian, 2019). A student may also be subject to a lower quality of peers at a less selective university, which again may have a negative impact on their performance (Winston and Zimmerman, 2004). Finally, if employers value the reputation of university attended, graduates of less selective universities may have less success in the labour market, regardless of their university experience (Walker and Zhu, 2013). Mismatch therefore has important implications for equality, social mobility and the life chances of those from disadvantaged backgrounds.

Despite the growing research base in the US, there is a paucity of research on mismatch in the UK context. Research by the Sutton Trust (2004) revealed that only 26% of state school students achieving 3 good A levels went on to a leading university, compared to 45% from independent schools. Related research (Chowdry et al, 2013) shows that disadvantaged students are less likely to attend Russell Group institutions, even when they are high achieving. However, to date, there has been no UK research examining the extent of mismatch across all universities, what these students look like in terms of background characteristics, or whether mismatched students in the UK go on to have poorer academic and labour market outcomes. This research project fills this gap, by being the first to study the phenomenon of mismatch in UK higher education.

This project speaks to the current policy developments in social mobility and higher education. Under guidelines published in the HE White Paper (BIS, 2016), from 2017, universities are assessed according to their performance on student satisfaction, retention and graduate employment. Our work is therefore relevant to universities and policymakers who are seeking to understand why certain students are more at risk of low performance on these indicators, as well as providing potential solutions. The government also intends to make it easier to set up "high quality" universities in order to give students more choice. The entrance of new providers into the market makes it even more crucial that students make the correct choice of university. Understanding which students undermatch and the consequences will help policymakers to develop better information and guidance strategies to help students with these choices.

This research therefore sheds new light on an un-researched issue, and presents new evidence on this potential driver of the socio-economic gap in degree and labour market success. In doing so it contributes to policy on widening participation, social mobility and potentially on the design of the higher education application system. It also has implications for information, advice and guidance strategies.

The remainder of this report proceeds as follows. Section 3 explains the institutional context while Section 4 describes the data used in the project. In Section 5 we describe the methodology used to identify mismatched students, and to understand their characteristics and outcomes. In Section 6 we describe the extent of student to university mismatch in the UK higher education sector. Section 7 builds on this by describing the characteristics of mismatched students. In Sections 8 and 9 we examine the severity of mismatch and the key drivers of SES and gender gaps in mismatch. We then illustrate the impact of being mismatched on degree outcomes (Section 10) and labour market outcomes (Section 11). Section 12 concludes.

3 Background

We are interested in the university and subject choices made by students from England. There are particular key features of the English education system which are relevant to how students make these choices, and in the potential barriers to match they may face.

At a relatively early age (13/14), students choose the types of qualifications and, crucially, subjects that they will study in their final two years of compulsory schooling, most often for 8 to 10 General Certificates of Secondary Education (GCSEs). Those who stay on after the compulsory schooling age[6] face another set of important decisions regarding their qualification and subject choices from age 16 to 18, most commonly in the form of 3 A levels. Finally, students wanting to study for Bachelor's degrees then have to choose both an institution and subject (course) to study for their higher education. Students apply to up to five courses using a centralised system (UCAS) and universities make offers based on their *predicted* grades in their top 3 A levels. Students then choose which offer to accept and, conditional on achieving the grade combination offered, attend their chosen university course.

Thus, there are a number of elements of this process that could give rise to mismatch, particularly if students are poorly informed about the process.

First, students could choose an A level subject combination that turns out to be unhelpful for gaining access to their preferred university. Research (Dilnot, 2016) has noted that there are socio-economic gaps in the types of A levels taken by students, and that these choices could impact the quality of institution they can ultimately access. For example, the more elite universities such as those in the Russell group favour academic subjects, but low SES students are less likely to take these subjects, and more likely to take vocational A levels such as law, which are less helpful for university entry. Although the Russell Group provide advice on the kind of subject choices that will give young people "the most options" when it comes to accessing high quality institutions, students from low SES backgrounds may not realise that taking subjects that are not on this list could hinder them.

Second, students apply to universities based on their predicted, rather than their actual grades (an issue that is subject of current political debate). Research (Wyness, 2016) has shown that

[6] We study the cohort who left school in June 2008. The compulsory schooling age at this time was 16. It is currently still 16, though since 2015 pupils have been required to stay in full-time education, for example at a college, start an apprenticeship or traineeship, or spend 20 hours or more a week working or volunteering, while in part-time education or training until age 18. See https://www.gov.uk/know-when-you-can-leave-school for more details.

whilst low SES students tend to be more likely to have their grades over-predicted (thus encouraging them to apply to high status institutions), high-attaining applicants from low income backgrounds are significantly more likely to have their grades under-predicted than those from high-income backgrounds. This is important because under-predicted candidates were also more likely to apply to, and to be accepted to a university which they were overqualified for – i.e. to undermatch. While we cannot examine the impact of grade predictions in this report, due to lack of available information, this is a potential mechanism, and an avenue for future research.

In summary, the English system is highly complex, and compels students to make decisions very early on in their school careers. This could be an issue if certain types of students (e.g. those from low SES backgrounds) lack information that could help them make the best decision. Indeed information problems have been commonly cited as an important driver of mismatch in the US setting (Hoxby and Avery, 2012; Dillon and Smith, 2017).

However, there are also some important advantages of the English system in terms of enabling students to choose an institution that matches to their attainment. Unlike students in the US, English students face far fewer financial barriers. In the cohort we study (students leaving school and entering university in 2008/09), tuition fees were £3000 per year (nominal prices), and backed up by an income-contingent loan. Further loans were available to all students for maintenance costs, and poorer students could also access a grant of around £2,800 per year nominal. This means there were no upfront costs, and students could afford to support themselves through university (Murphy et. al. 2019). Another important feature of the UK system is that there was little price variation between institutions (or subjects) meaning that students could not be led to choose a cheaper university or course (which may be a worse match) for financial reasons. For example, for our cohort of interest, Cambridge University (ranked 1 in UK) charged £3,000 per year for its courses, while Bolton University (ranked in 125th UK) also charged £3,000 per year.

Thus, it is ambiguous as to whether we would expect to see a lesser or greater degree of mismatch in England vs the US (where, to date, all the existing research on mismatch has taken place). However it is important to note that the English context is quite different than the US. Drop out is considerably lower in England than in the US. According to HESA (2019), drop out rates in the UK stand at 6.3% for young first degree entrants in 2016/17. As a point of comparison for the US, the six-year graduation rate for students at public four-year colleges is around 60.5%, and at private non-profit colleges is 62.5% (Shapiro et al, 2014). Completion rates are often used as the main measure of degree performance in the US, making comparison based on degree performance between the two countries tricky.

4 Data

In order to identify mismatched students, we use individual-level administrative data on the population of state-school students in England for a single cohort. Our focus is on the cohort of young people who took their compulsory age 16 exams in 2006. They will have taken their non-compulsory end of secondary school exams two years later in 2008. The grades from exams are used to determine which university a student will be admitted to. The students then enter university in the autumn of that year at age 18 (the traditional age for university entry in England) or 19 if they took a gap-year (around 25% of our sample – see Table 1). Our data cover students in all publicly-funded English schools[7], and we combine this with information on the university course attended by these students anywhere within the UK (England, Scotland, Wales, and Northern Ireland). We also incorporate aggregated data on the earnings outcomes of an earlier university cohort, which are based on tax records.

Our schools data come from the National Pupil Database (NPD), and include basic demographic information (gender, ethnicity, English as an additional language, special educational needs) alongside exam results at ages 11, 16, and 18. There is substantial attrition over this period of education in the English system from age 16 (the end of the compulsory school period in the period we study). At this point, pupils either go on to 6th form college, or leave the schooling system altogether, some then moving to further education colleges to study for below-degree/technical qualifications and some going into employment, or neither (in all, 60% of our cohort leave the schooling system at age 16). Of those continuing on to 6th form college, a smaller group leave school/6th form college at age 18 without going on to university (around 10% of our cohort). Our main interest is in the subgroup who go on to university, but we use information on the complete population of age 16 students to construct key variables, as we describe below. Starting with a sample of around 590,000 in the 2006 cohort, we initially restrict the sample to all university students who went to a state-school, and on whom we have information on demographics at age 16 and exam results at age 18, which results in a sample of 138,535.

[7] 93% of students attend publicly funded secondary schools in England (DfE, 2010)

Our linked data on course attended[8] come from the Higher Education Statistics Agency (HESA). We use university entry information from 2008 and 2009, since a substantial proportion of students in England delay university entry for one year after age 18 examinations. These data contain information on every student's course in every higher education establishment in the UK. We have access to four digit JACS (Joint Academic Coding System) codes, which separately classify around 1,300 different university subjects. For example, we can separately identify those who are studying Economics from those who are studying Applied Economics, and those who are studying History by period from those who study History by topic.

Using data from 2008 and 2009 is beneficial, since we can observe the labour market outcomes of these students, which would, of course, be impossible with later cohorts. However an obvious downside is that our data is less recent. The number of students going to university has increased since 2008, and there has been increased activity devoted to widening participation. However, the gap between rich and poor students has remained relatively stable at best, depending on the definition used (Murphy et al, 2019) suggesting our findings are reasonably representative of the current situation.

Our aggregated earnings data come from the new Longitudinal Education Outcomes (LEO) dataset. They are compiled from tax records by Her Majesties Revenue and Customs (HMRC) in the UK, and are publically available in the aggregated form in which we use them. We use the median earnings outcomes five years after graduation for the earliest available cohort, which is those who completed undergraduate degrees in 2009. These data are available for 23 broad subject categories at each university, so they are more aggregated than the subject information we have for students. To ensure comparability between our match measures, we collapse the 1,300 university subject categories into these broader groups.[9]

4.1 Measuring socio-economic status

To construct a measure of students' socio-economic status we follow Chowdry et. al. (2013). This method uses information on whether a student was eligible for free school meals at age 16, alongside a set of variables which describe the neighbourhood in which they live at that age. The free school meals measure is essentially an indicator of whether a student is from a household in receipt of state benefits (15% of students). In order to capture a broader set of

[8] Note that as in Dillion & Smith (2017) we observe a collapsed version of the student-course match process, in that we only observe the course that they attend, rather than where they apply.

[9] Estimations using the finer measure of subject studied produces very similar results.

socio-economic circumstances, use a set of neighbourhood variables taken from the 2001 Census. These measures are available at the Lower Super Output Area level, which is a neighbourhood containing around 700 households or around 1,500 individuals. This includes the proportion of individuals in the neighbourhood that: 1) work in managerial or professional occupations ; 2) hold an A level equivalent qualification or above; and 3) own their home. In addition, we also use the derived ONS Area Classification (2001) and the 2007 Index of Multiple Deprivation.

Table 1 Summary Statistics

	Quintile of SES					
	1	2	3	4	5	Total
Student Characteristics						
Female	0.59	0.58	0.57	0.56	0.54	0.56
	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)
Took gap year	0.22	0.23	0.24	0.25	0.28	0.25
	(0.41)	(0.42)	(0.42)	(0.43)	(0.45)	(0.43)
Special educational needs	0.06	0.05	0.04	0.04	0.04	0.03
	(0.26)	(0.24)	(0.23)	(0.21)	(0.22)	(0.18)
English as an additional language	0.32	0.24	0.16	0.08	0.07	0.12
	(0.51)	(0.46)	(0.39)	(0.30)	(0.28)	(0.32)
Ethnic minority	0.41	0.35	0.25	0.16	0.14	0.18
	(0.53)	(0.52)	(0.48)	(0.42)	(0.42)	(0.38)
Sample composition						
Proportion of sample	0.08	0.14	0.19	0.24	0.34	1.00
N	11690	19843	26470	33413	47119	138535

We combine these measures using principle components analysis to create a standardised index.[10] We use the whole population of state-school students at age 16 in the relevant cohort to construct the index, so throughout this paper SES refers to socio-economic position relative to the whole school-cohort population rather than relative to the university-attending sub-population. The final row of Table 1 illustrates that this results in 8% of our sample coming from the most disadvantaged families and 34% from the least disadvantaged families.

[10] See Appendix Figure A1 for a discussion of how this measure compares to an alternative measure of parental socio-economic status.

Table 1 highlights the key characteristics of our sample by SES quintile. Women are overrepresented in higher education, making up 56% of the sample. Interestingly this is particularly true for lower SES families with women making up 59% of the most deprived students. A quarter of our sample took a gap year, with the least deprived families more likely to take a year out than the most deprived. There are only a small proportion of people with special educational needs in our sample as might be expected, with 6% of the most deprived families and 4% of the least deprived families being categorized in this way. Finally, there is a strong association between having English as an additional language, or being in an ethnic minority and being low SES with these groups accounting for a larger proportion of low SES families.

4.2 Aspirations and Preferences of mismatched students

In order to understand the aspirations and preferences of mismatched students, we use linked Next Steps-NPD data on students from the same cohort. Next Steps is a longitudinal survey of young people following the cohort who were aged 14/15 in 2004 and who would have first attended university in 2008/09 at age 18/19. The advantage of using this data is that it contains questions on students' HE attitudes, beliefs and choices towards further and higher education, allowing us to potentially understand the reasons why some students may mismatch.

In particular we examine the following questions:

Attitudes/Beliefs about getting into university (asked at age 15/16/ Next Steps wave 3):

- Whether student thinks they will apply to university
- Whether student believes they will get into university

Preferences around prestige and location (asked at age 17/18/wave 5):

- Whether location of the university is important
- Whether the prestige of the university is important

Whether HE subject/institution was first choice (asked at age 18/19/wave 6):

- Whether the university student is attending was not their first choice
- Whether the subject student is studying was not their first choice

Advice and guidance (asked at age 18/19/wave 6)

- Whether student spoke with teachers about university plans

4.3 Measuring undermatch in Next Steps

Given that Next Steps is linked to NPD data, this means we can, in theory, construct our measures of mismatch in this dataset. However, the Next Steps sample is small, with only around 10,000 young people surveyed, and, once restricted to those who attend university, with valid university and course choices, and linked NPD data, this sample shrinks to around 2,500 students. This means that there are many courses where only a few students are studying. This will be problematic for calculating our course quality measures. In order to resolve this issue, we impute information on our course quality measures (both for points-based and earnings-based) from the population administrative data (see p6 and Figure 1 for the detailed description of how these are constructed), assigning each course the same percentile as in our NPD-HESA data. For comparability, we also match in each student's percentile in the sample, assigning their percentile in the population-level data, based on their A level and equivalent test scores from the restricted NPD-Next Steps sample. We are therefore ensuring that our sample is allocated a course and individual quality measure that is representative of the whole population, minimising any impact of sampling bias and attrition. Reassuringly, the proportions of undermatched and over matched students are similar to that of our main administrative data (particularly for undermatched students). A comparison of the NPD-Next Steps sample versus the wider population data is presented in Section 8, alongside results.

4.4 Measuring SES in Next Steps

Unlike in our administrative sample, we are not able to create a composite measure of socio-economic status, since our data does not contain the necessary indicators for all students. Instead, we measure SES using parental national statistics socio-economic classification (NS-SEC) (taken at age 13), as follows:

Table 2: Family's NS-SEC class

	high SES	medium SES	low SES
Higher Managerial and professional occupations	x		
Lower managerial and professional occupations	x		
Intermediate occupations		x	
Small employers and own account workers		x	
Lower supervisory and technical occupations		x	
Semi-routine occupations			x
Routine occupations			x
Never worked/long term unemployed			x

4.5 Outcomes of mismatched students

Once we have identified mismatched students and their characteristics using the above datasets, we then go on to examine the outcomes of mismatched students. Here, again, we use linked individual-level administrative data (NPD-HESA) on the same population of students. To observe their later university and labour market outcomes we use a further link to more HESA information and to the Destination of Leavers of Higher Education (DLHE) survey, which provides us with the labour market outcomes of our cohort of 2008 attendees, 3.5 years after graduation, in 2014. This enables us to measure:

University outcomes:

- Whether the student dropped out of university
- Whether the student got a 'lower class degree' (less than a 2:1)

Labour market outcomes:

- Whether the student was not in employment or education 3.5 years after graduation
- The labour market earnings of the students (full-time employed students only).

Given that HESA only constructed a subsample[11] for the longitudinal element of the data, we explore the similarities in characteristics between our main sample and those who we observe later labour market outcomes for in Table 2. This illustrates that despite a large amount of attrition for those who we observe our 'lower class degree' and employment outcomes for, the demographic characteristics of the samples are very similar. The smaller sample for the earnings outcome is slightly less representative of the full sample, particularly for low attainers (bottom 20% of student performance at age 18), where there are more females and fewer ethnic minorities included. High attainers (top 20% of student performance at age 18) are more similar in terms of extent of undermatch, gender and SES, although there are slightly fewer ethnic minorities included.

[11] The process used to construct the subsample is explained here: https://www.hesa.ac.uk/data-and-analysis/publications/long-destinations-2012-13/notes (accessed 27/3/2019), 10.24am. Also see the technical report found here: https://www.hesa.ac.uk/files/LongDLHE_1213_TechnicalReport_HESA_IFF.pdf

Table 3: Key characteristics for different samples across outcomes

	University outcomes (HESA)		Labour market outcomes (DLHE)	
	Drop out (full sample)	'Lower class degree' (N=55,783)	Not in employment or education (N=47,606)	Labour market earnings (N=14,601)
Low attainers				
Points-based overmatch	34	31	31	34
Earnings-based overmatch	55	44	47	51
Female	49	55	53	57
Ethnic minority	23	21	20	14
High SES	68	68	69	70
Low SES	32	32	31	30
High attainers				
Points-based undermatch	24	25	26	25
Earnings-based overmatch	39	46	47	42
Female	55	58	58	60
Ethnic minority	17	17	16	14
High SES	81	80	80	80
Low SES	19	20	20	20

Notes: Mismatch is defined as 20 percentiles above or below zero for the purpose of this analysis. High and low SES are measured in HESA rather than NPD, capturing the parents' NS-SEC. High SES is those who report higher managerial and professional roles while low SES is those who report routine and manual roles.

5 Methodology

5.1 Measuring mismatch

We are interested in the match between student quality (attainment) and course quality. We measure student quality according to their academic attainment, based on their performance in their best 3 A level or equivalent exams at age 18[12]. Each student is given a percentile rank in

[12]We consider only the students who go on to university, so the relevant exam results distribution is that of university attendees. Some students take courses that are equivalents to A levels. In these cases we calculate their A level equivalence scores.

the national attainment distribution of exam scores of all English state-school students in that cohort[13]. A student ranked 100 is at the top of the distribution (with 3 As), and a student ranked 0 is at the bottom.

We create two measures of course quality:

Our first measure of course quality (points-based) is created based on the total of the best 3 exam scores at A level or equivalent (as above) of the *median student on that course*. The lowest ranked course is assigned a value of 0 and the highest a value of 100.

For our second measure of course quality (earnings-based), we rank each course based on the earnings of a previous cohort of graduates from that course, 5 years after graduation. The courses with the lowest average earnings are assigned a value of 0 and those with the highest average earnings are assigned a value of 100.

With this information, we create two match measures.

(i) **Points-based match**. We subtract the student's percentile in the exam results distribution from the percentile of their course on the course points-based quality distribution. Here a student would be well matched to their course if their academic attainment is roughly the same as that of their fellow students on the course. Students are defined as undermatched if they have higher academic attainment compared to their fellow students on the same course (negative match index), and are defined as overmatched (positive match index) if they have lower academic attainment compared with their fellow students on the same course.

(ii) **Earnings-based match**. We subtract the student's percentile in the exam results distribution from the percentile of their course on the course's earnings-based quality distribution. Here, a well-matched student would be one who is relatively high-attaining, and attends a course with high earnings, or is relatively low-attaining and attends a course with relatively low earnings. Students are then defined as undermatched (negative match index) if they are high-attaining, but attend a low earning course, and students are defined as overmatched (positive match index) if they are low-attaining but attend a high earning course.

[13] As distinctive feature of the English education system is the importance of subject choices made in secondary education and at university and we take this into account by adjusting for subject difficulty following Coe et al. (2008). See Appendix for details of this process.

Both measures of match reflect different aspects of course quality, and the measures can produce quite different relative positions for the same course. For example, a course which is positioned towards the bottom of the point score distribution, Computer Science at Southampton Solent (commanding a median difficulty-adjusted points score of 408), is considered high quality in terms of earnings, ranked at the 70th percentile, with average earnings of around £27,500. In contrast, English at Edinburgh is ranked at the 90th percentile on our points-based quality measure (median difficulty-adjusted points score of 801), but is only ranked at the 35th percentile in terms of our earnings-based measure (average earnings of £23,300).

5.2 Extent of mismatch

Our first aim is to understand the extent of mismatch. Here we present both summary statistics of the proportion of students who mismatch, and graphical results describing the extent of mismatch among the population.

Since low-attaining students are more likely to overmatch for mechanical reasons (such students are at the bottom of the attainment distribution, therefore are mechanically forced to attend higher quality courses), and high-attaining students are more likely to undermatch for the same reasons, throughout our analysis we present our results separately for the bottom and top attainment quintiles.

5.3 Who are mismatched?

Our second aim is to understand whether there are inequalities in match. In particular, we might have reason to believe that low socio-economic status (SES) students maybe more likely to undermatch than their counterparts from richer backgrounds, or that women may be more likely to undermatch than men. To analyse inequalities in match, we run multivariate regressions with our match index as the outcome variable, controlling for background characteristics (SES, gender, ethnicity, English as an Additional Language (EAL), and statemented and non-statemented SEN), prior academic attainment at age 11 and 16, and school attended. We present the results of these regressions graphically.

We are also interested in those students who severely under or overmatch – e.g. a very high-attaining student who attends a very low quality university, or vice versa. Here, we are interested in whether, again, there are SES or gender gaps among these students at the extremes. We illustrate gender and SES gaps for these extremely undermatched and extremely overmatched students.

5.4 Drivers of mismatch

To try and understand the drivers of university mismatch, we again use multivariate regressions to consider the potential drivers of SES and gender gaps. To this end, in our regressions we (separately) include controls for:

- subject choice
- geographical factors (distance to university attended, and distance to each of the nearest 3 universities to the student's home neighbourhood, and distance to all remaining universities)
- school-level factors (the average SES index of pupils in the school, the proportion from the school attending university, and the school itself).

5.5 Outcomes associated with mismatch

When we consider the association between mismatch and university and labour market outcomes (sections 9 and 10), we use multivariate regressions for each of our four outcomes. Our main explanatory variable is defined as those who are 20 percentiles above or below zero (matched) for low (overmatched) and high (undermatched) attainers[14]. To compare similar individuals, we add controls into the models in stages controlling for:

- Demographics including SES, gender, ethnicity, English as additional language, Special Educational Needs, region of residence.
- School factors, including the school itself
- University group: Russell Group vs non-Russell Group.

We also present some sub-group analysis, based on the broad subject grouping that the student studied at university (STEM, social science, arts and humanities), to consider how subject choices might be driving the differences in outcomes.

6 What is the extent of mismatch in the UK system?

Figure 2 shows the distribution of our two measures of student-course match – points-based match and earnings match. Those students with negative match indices are undermatched (attending a course that is lower quality than expected, given their grades) and students with

[14] This approach allows us to separate out differences in outcomes for two distinct groups of mismatched students, those who undermatch compared to those who match, and those who overmatch compared to those who match, keeping our baseline comparison group (matched students) consistent across models.

positive indices are overmatched (attending a course that is higher quality than expected, given their grades).

Two things are apparent from this chart.

- First, for both types of match, the distributions are approximately symmetrical, with equal proportions of students undermatched and overmatched.
- Second, the earnings-based measure is more spread out than the points-based measure. This is likely because students choose courses on the basis of their academic points, thus we would expect students to be better matched to their course on this measure. Meanwhile, students may not be well-informed of the potential earnings of each course as this information is not widely known to the public.

Following Dillon and Smith (2017), we can calculate the proportion of students undermatched and overmatched in our sample. Dillon & Smith define mismatch as:

- Student is undermatched if they are 20 or more percentiles below the position of their course
- Student is matched if they are within 20 percentiles of their course
- Student is overmatched if they are 20 or more percentiles above the position of their course

Given this measure, our results are as followed (compared with Dillon and Smith, for interest)

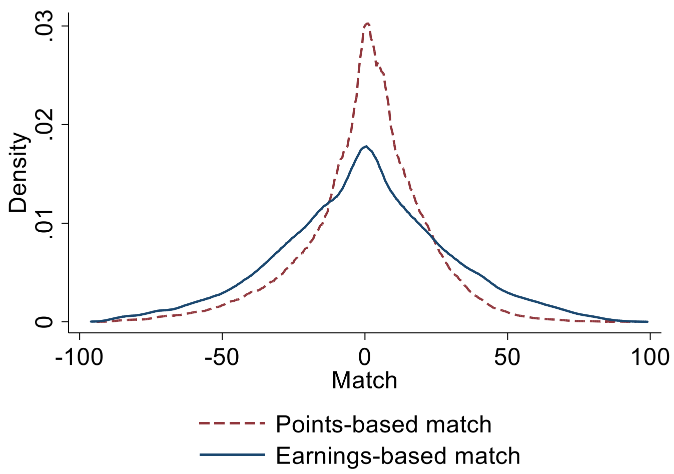
Table 4: The extent of mismatch, and comparison with Dillon and Smith

Measure	Undermatched %	Matched %	Overmatched %
Dillon and Smith	25	50	25
Our points-based measure	15	70	15
Our earnings-based measure	23	54	23

It would be problematic to draw strong conclusions from these comparisons, given the underlying differences in the measures. In particular, Dillon and Smith's measure of university quality is a composite measure, based on four elements: mean SAT score (or mean ACT score converted to the SAT scale) of entering students, percent of applicants rejected, the average salary of all faculty engaged in instruction, and the faculty-student ratio. This is quite a different measure from both our points-based and earnings-based quality measures. None-the-less, this provides a simple benchmark by which to characterise the extent of mismatch within the English system.

In the analysis that follows in Sections 7-9, we prefer to use our continuous measure as it provides information about the distribution of mismatch and allows for estimation of the impacts at the extremes of mismatch. However, we revert to using this categorical measure of match in Section 10 and 11, where we examine match outcomes, for clarity purposes.

Figure 2: Points-based and earnings-based measures of student-course match



Source: NPD-HESA and LEO. n= 138,535. University attendees only.

7 Who are the mismatched students?

In this section, we focus on SES and gender gaps in mismatch as our main inequalities of interest. We also examine ethnic differences in match.

Figure 3a shows a simple plot of students achievement (x-axis) against points-based average course quality (y-axis) for all students at each position in the attainment distribution (though we aggregate into bands of 5 percentiles, for clarity). Figure 3b is the same plot, but for earnings match.

If all students were to match to their courses – that is, to have the same qualifications as the median student on their course (for points match) or to be at the same position in the attainment distribution as their course is on the earnings distribution (for earnings match) this line would be straight and at a 45-degree angle. The extent to which the line is above the 45-degree line indicates how overmatched these students are on average, and similarly how far below the 45-

degree line reveals the extent of undermatch. Plotting this match-line for different types of students allows us to study the match gap at any point in the attainment distribution.

These line charts highlight the following:

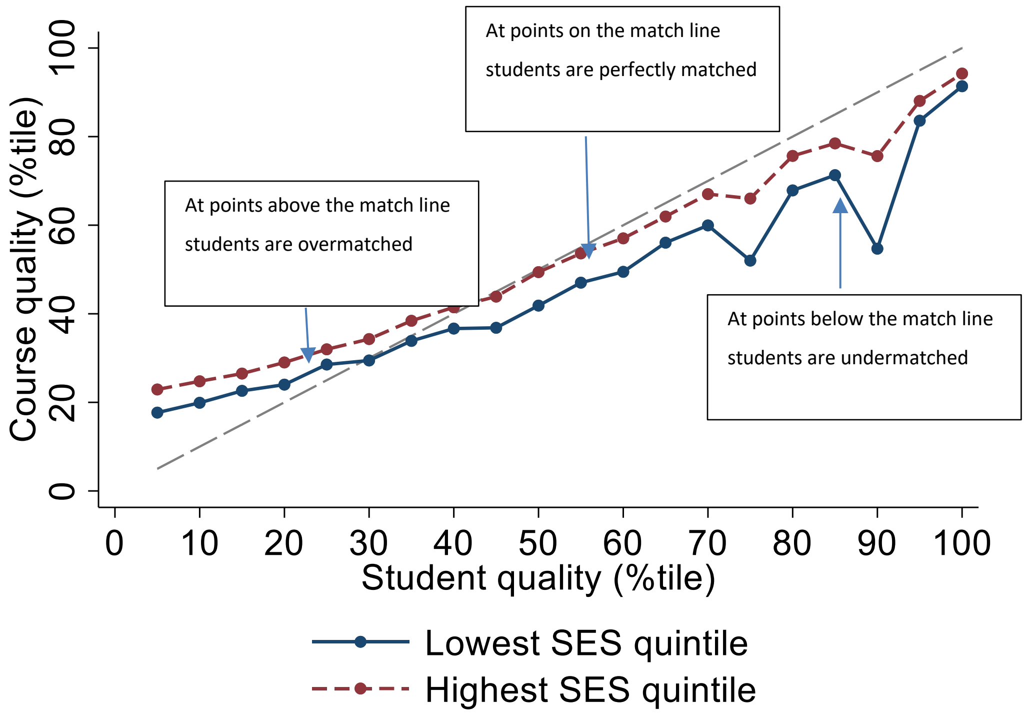
Points-based match

- Low attainers are more likely to be found above the 45 degree line – in other words, they are more likely to overmatch. Meanwhile, high attainers are more likely to be found below the 45 degree line, being more likely to undermatch. As previously explained, this is expected, and largely mechanical (as described on p28).
- There are large SES gaps in match quality. For every given level of individual attainment, high SES pupils attend higher ranked courses than similar attaining low SES pupils. This means that pupils from disadvantaged backgrounds are more likely to be undermatched, and less likely to be overmatched. Thus, it would seem that high SES students pursue course quality to a greater extent than low SES students, regardless of attainment.

Earnings-based match

- Again, we see low attainers being more likely to overmatch and high attainers more likely to undermatch. But we also see that there is a higher degree of mismatch on this measure compared to the points-based measure. As previously explained, this is likely to be because students are accepted to universities on the basis of their points scores, so we would expect to see a higher correspondence between student and university quality in the points-based measure.
- There are also large SES gaps in earnings match quality. Indeed, for every given level of individual attainment, those from high SES backgrounds attend courses with higher earnings than low SES pupils. Thus, it would seem that high SES students pursue higher earning courses to a greater extent than low SES students, regardless of their attainment.

Figure 3a: Line graphs, SES, attainment match



Next, we present the same mismatch charts but for gender gaps in match. These are shown in Figures 4a and 4b. These figures reveal the following:

- There are almost no gender gaps in points-based match, apart from at the very top of the attainment distribution, where high-attaining females chose courses that are slightly less academically selective than men of similar attainment levels.
- However, there is a large gender gap in earnings match. Here, women are clearly choosing lower earning courses than men. It is interesting to note that from the above we can see that women are choosing courses that are as academically selective as men, but across the entire attainment distribution, they choose courses that attract lower earnings.

Figure 3b: Line graphs, SES, earnings match

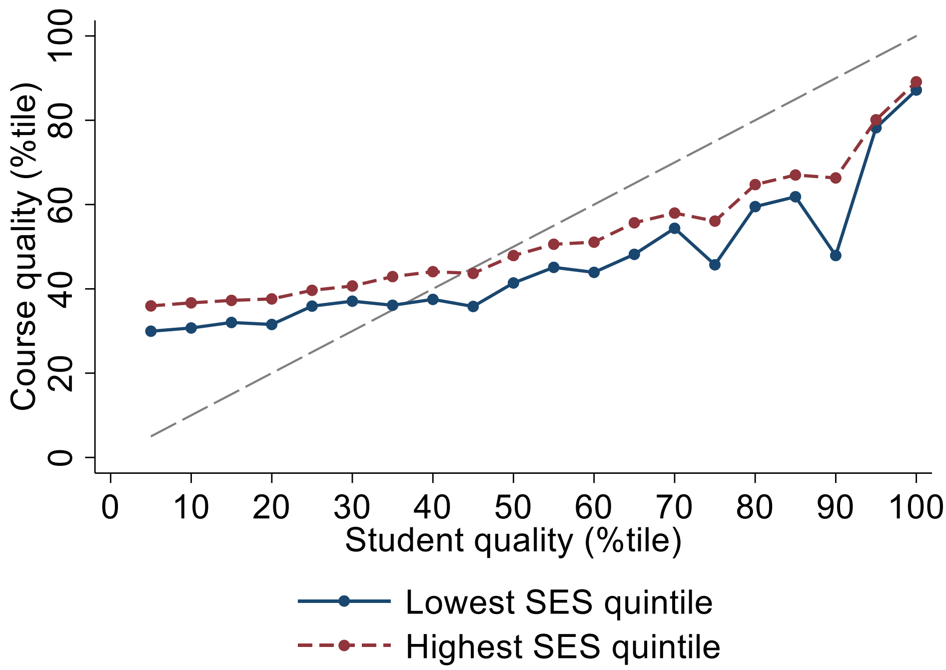


Figure 4a: Attainment match by Gender

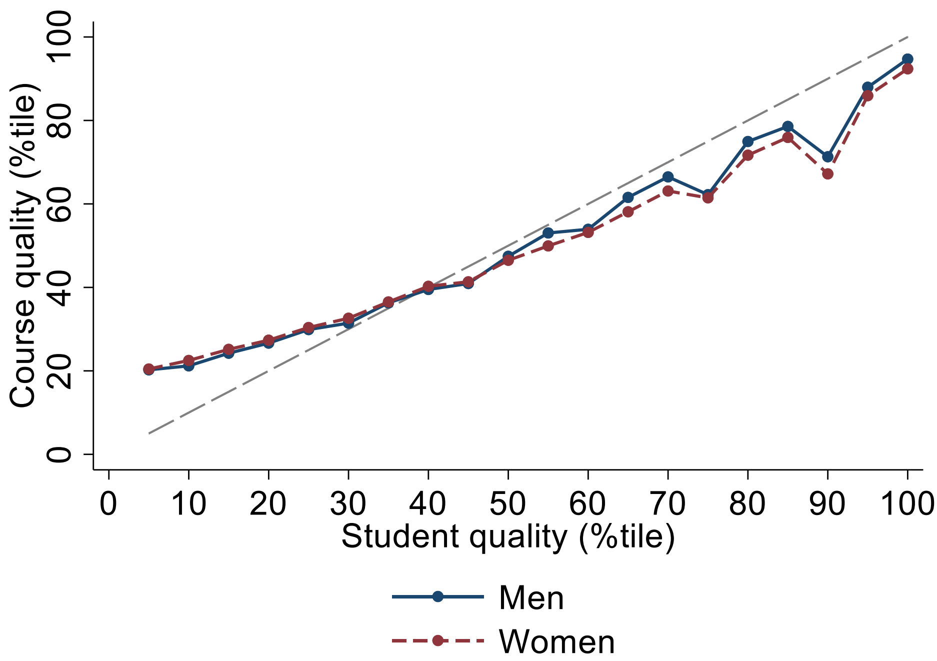
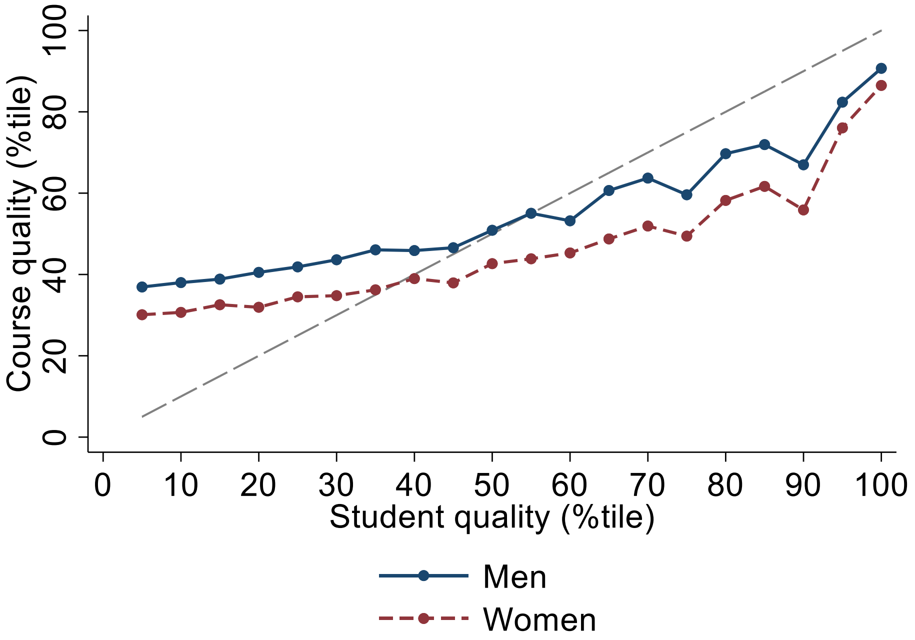


Figure 4b: Earnings match by Gender



Next, we wish to test whether these SES and gender gaps in match are statistically significant, and if they persist once we control for background characteristics and prior academic attainment.

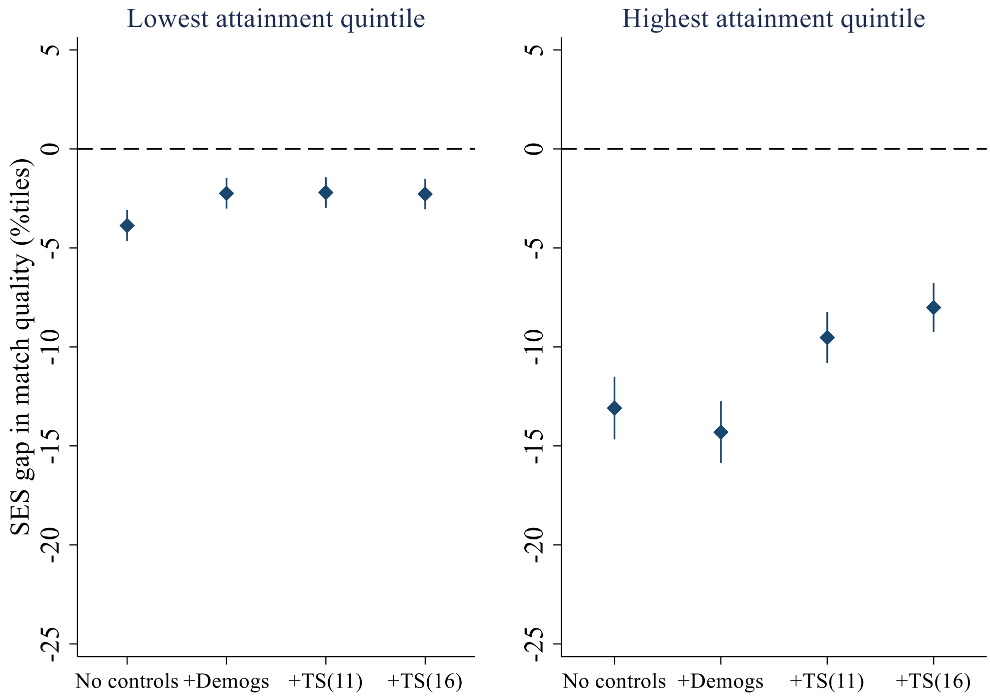
In Figures 5 – 6 we therefore present regression coefficients from our multivariate models in which we explore the relationship between match quality and SES/gender, conditional on student characteristics and prior attainment up until age 16. This shows the gaps in match quality for the lowest SES quintile relative to the highest (Figures 5a and 5b), and women relative to men (Figures 6a and 6b). We also include models of ethnic grouping relative to White (Figures 7a and 7b). The models are estimated separately for the high and low attainment groups (top 20% and bottom 20% of attainers) and for each of the match measures.

These charts reveal the following:

SES gaps – points-based (Figure 5a)

- Among low attainers (who, as previously established, typically overmatch), there is a significant, though small raw SES gap; those from disadvantaged backgrounds are 5 percentiles less overmatched than those from advantaged backgrounds (first column of figure). This corresponds with the gap in the match quality index seen in Figure 3a, for low attainers. After controlling for individual characteristics (2nd column), and prior attainment at age 11 (3rd col) and age 16 (4th col), this gap is slightly reduced, but persists, so that low attainers from disadvantaged backgrounds are 2percentiles less overmatched than high SES low attainers, conditional on these variables.

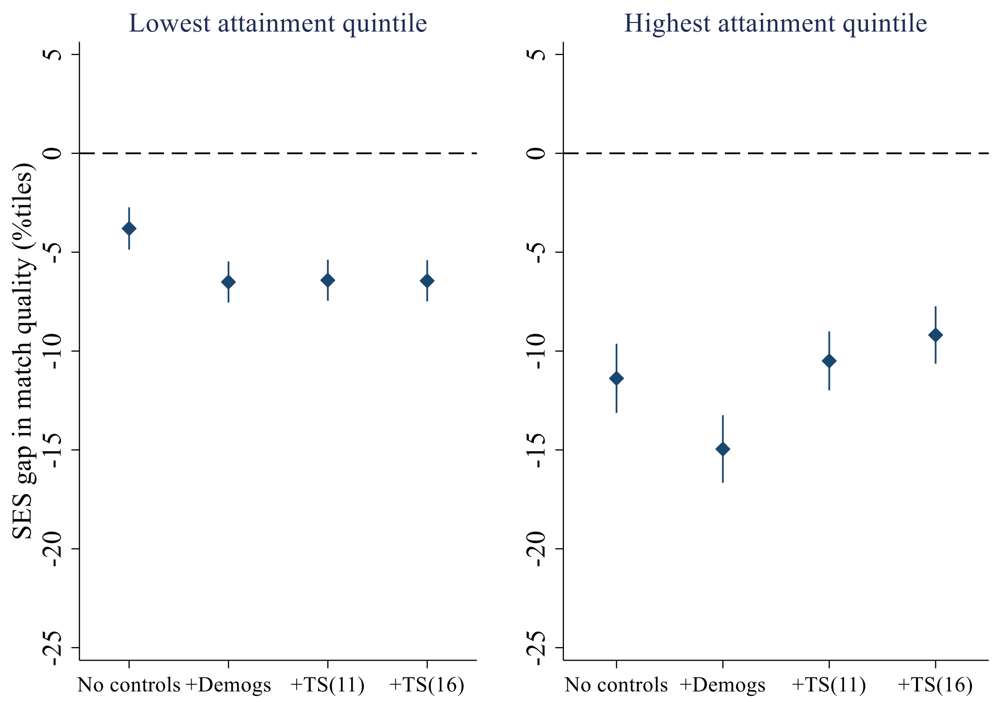
Figure 5a: SES gap (low SES-high SES) in points-based match with increasing controls



Notes on how to read these charts: the points on the chart show the socio-economic gap in match (low SES-high SES) in percentiles. A coefficient of zero would mean there is no SES gap, so that both low and high SES students choose the same quality of courses. A negative coefficient means that low SES students are lower matched than high SES students, choosing courses which are lower quality, for a given level of attainment. For example, a coefficient of -5 means that low SES students attend courses which are 5 percentiles lower quality than high SES students from the same attainment quintile.
Demogs: demographics, TS(11): test scores at age 11; TS(16): test scores at age 16

- Among high attainers (who typically undermatch), the raw SES gap is significantly larger – disadvantaged low attainers undermatch by 15 percentiles more than those from advantaged backgrounds. Again, this raw gap can be seen directly in Figure 3a, for those at the top of the attainment distribution. This gap is largely unchanged with the addition of student characteristics, but is substantially reduced once attainment at age 11 is taken into account. In other words, within the subset of individuals with similar age 11 attainment, the SES gap in match is narrower. Taking prior attainment into account therefore reduces the extent of undermatch. Nevertheless, the SES gap remains significant, confirming that, conditional on pupil characteristics and prior attainment, high-attaining low SES pupils undermatch to a greater extent than high-attaining high SES pupils. This 8-9 percentile gap corresponds to the difference between studying economics at the London School of Economics (ranked 5th in the Times Higher UK university rankings) versus Exeter (ranked 18th). This could have real labour market consequences for the student; the median earnings difference five years after graduation between these two courses is £13,200 per year.

Figure 5b: SES gap (low SES-high SES) in earnings-based match with increasing controls



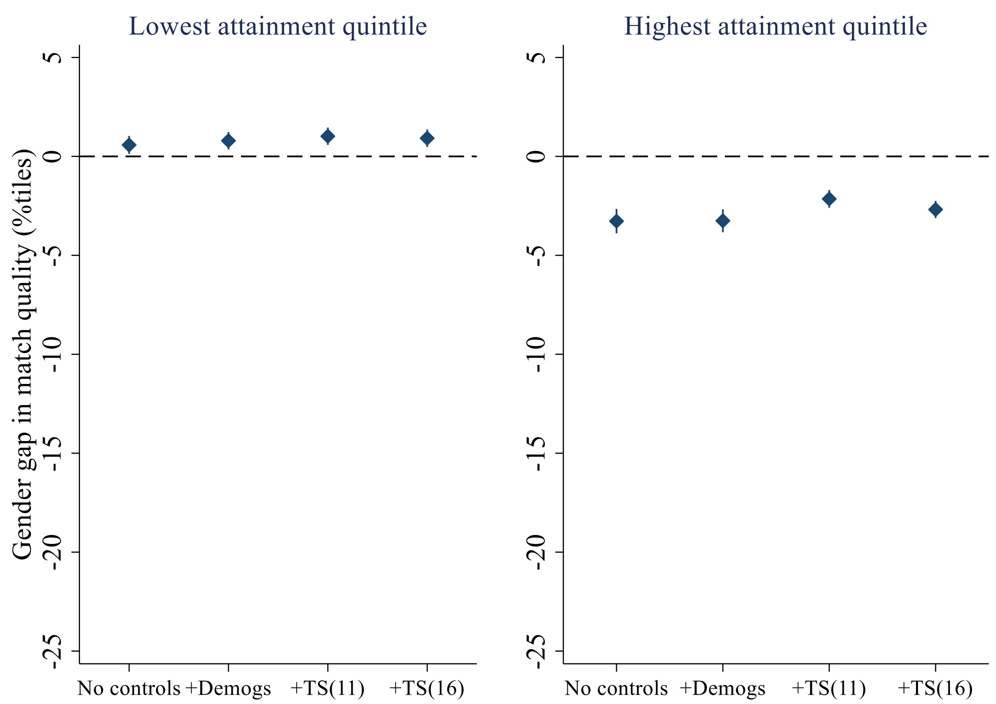
Notes on how to read these charts: the points on the chart show the socio-economic gap in match (low SES-high SES) in percentiles. A coefficient of zero would mean there is no SES gap, so that both low and high SES students choose the same quality of courses. A negative coefficient means that low SES students are lower matched than high SES students, choosing courses which are lower quality, for a given level of attainment. For example, a coefficient of -5 means that low SES students attend courses which are 5 percentiles lower quality than high SES students from the same attainment quintile.
Demogs: demographics, TS(11): test scores at age 11; TS(16): test scores at age 16

SES gaps – earnings-based (Figure 5b)

- The SES gap in earnings match quality is far higher than for our points measure of match, standing at 6 percentiles, even after controlling for characteristics and prior attainment. This suggests that even among the lowest attainers, high SES pupils attend courses with significantly higher labour market rewards.

- This is suggestive that low-attaining pupils from high SES backgrounds are more familiar with the earnings returns to certain courses, and value this characteristic when choosing a university.
- Among high attainers, the patterns are more similar to the points-based measure. Put simply, high-attaining students from advantaged backgrounds attend courses associated with higher earnings than their disadvantaged counterparts, even accounting for characteristics and prior attainment.

Figure 6a: gender gap (female-male) in points-based match with increasing controls

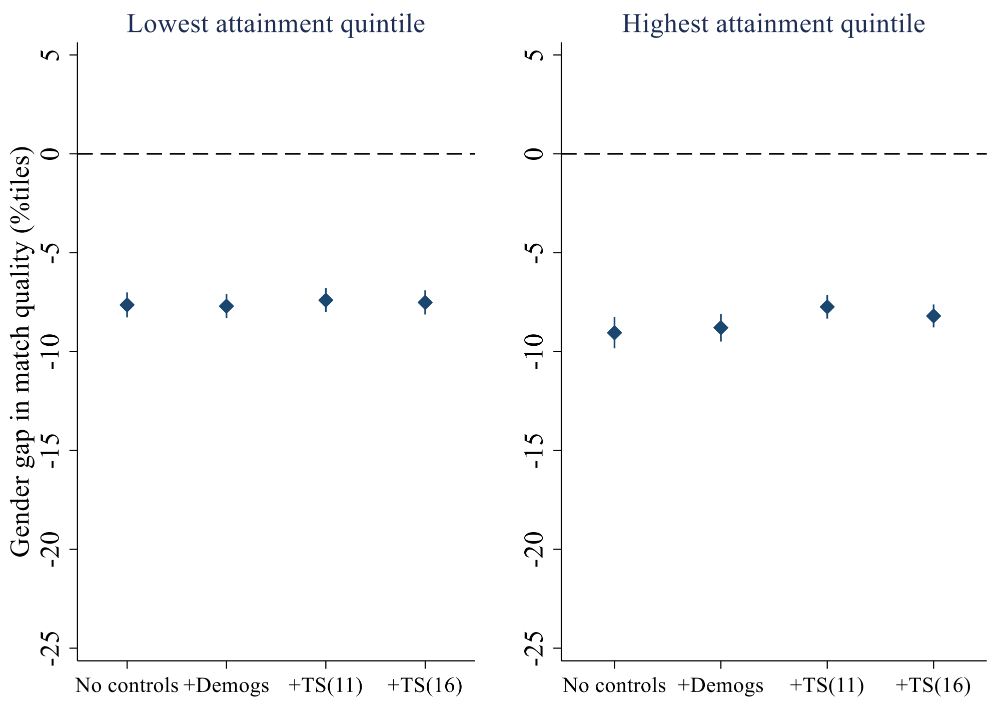


Notes on how to read these charts: the points on the chart show the gender gap in match (women-men) in percentiles. A coefficient of zero would mean there is no gender gap, so that both male and female students choose the same quality of courses. A negative coefficient means that women are lower matched than men, choosing courses which are lower quality, for a given level of attainment. A positive coefficient means that women are higher matched than men, choosing courses which are higher quality, for a given level of attainment. For example, a coefficient of -5 means that women attend courses which are 5 percentiles lower quality than men from the same attainment quintile.
Demogs: demographics, TS(11): test scores at age 11; TS(16): test scores at age 16

Gender gaps – points-based (Figure 6a)

- Figure 6a provides evidence on the gender gaps in points-based match among low and high attainers. These charts reveal little evidence of gender gaps in this dimension. For low attainers, there is effectively no gender gap; men and women choose courses that are equally academically selective, conditional on controls. For high attainers, there is evidence of a small gender gap in favour of men; high-attaining women attend courses that are slightly less academically selective than men.

Figure 6b: Gender gap (women-men) - earnings-based match with increasing controls

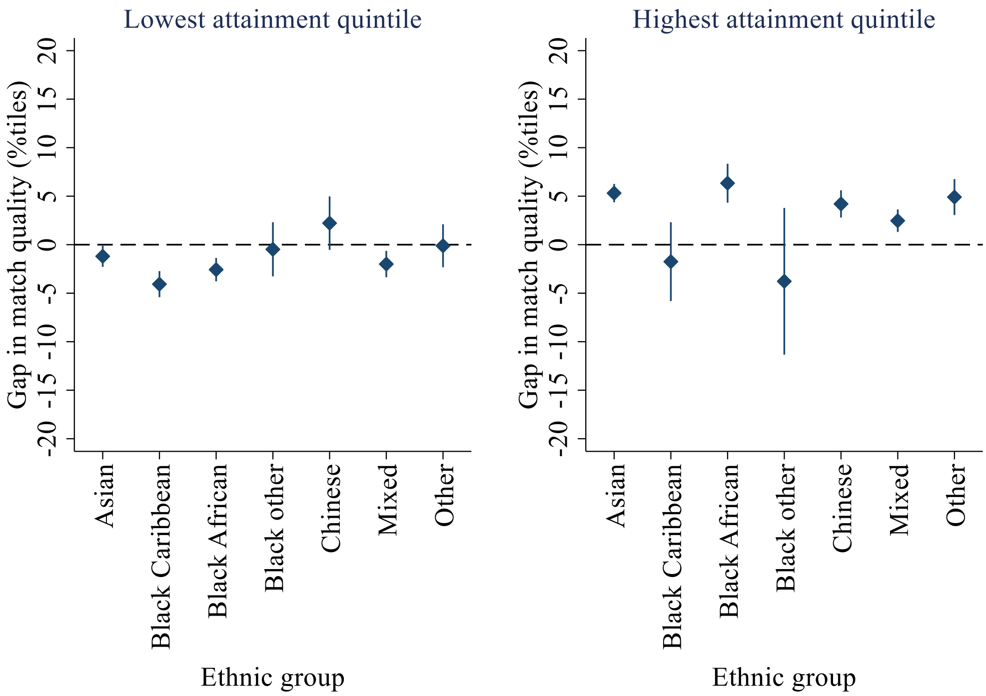


Notes on how to read these charts: the points on the chart show the gender gap in match (women-men) in percentiles. A coefficient of zero would mean there is no gender gap, so that both male and female students choose the same quality of courses. A negative coefficient means that women are lower matched than men, choosing courses which are lower quality, for a given level of attainment. A positive coefficient means that women are higher matched than men, choosing courses which are higher quality, for a given level of attainment. For example, a coefficient of -5 means that women attend courses which are 5 percentiles lower quality than men from the same attainment quintile.
Demogs: demographics, TS(11): test scores at age 11; TS(16): test scores at age 16

Gender gaps – earnings-based (Figure 6b)

- But the evidence in Figure 6b presents clear evidence of large and statistically significant gender gaps in earnings match. The raw gaps confirm those seen in Figures 4a and 4b; while women choose courses that are similarly academic as men, they choose courses that produce lower returns. This result holds after the inclusion of controls, with the gender gap reaching 10 percentiles for high attainers.
- This is a key finding of our research – that for a given level of academic attainment, women attend courses that are almost as academically selective as men, but with substantially lower earnings potential.

Figure 7a: Ethnic gaps (ethnic minority-white) in match points-based

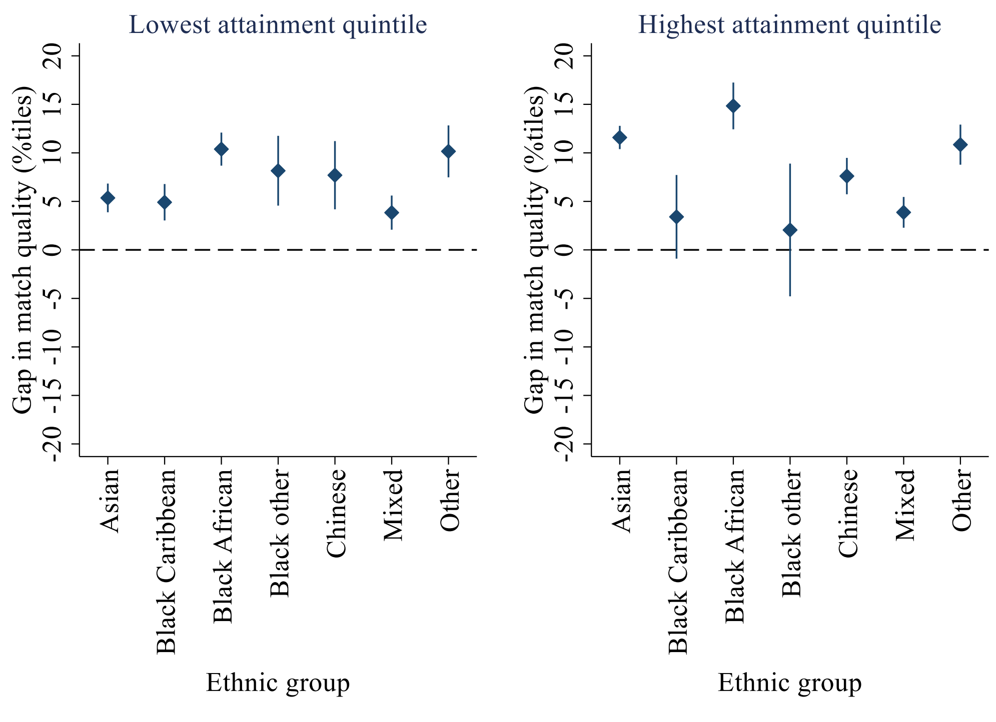


Notes on how to read these charts: the points on the chart show the ethnic gap in match (ethnic minority-White) in percentiles. A coefficient of zero would mean there is no ethnic gap, so that both ethnic minority and White students choose the same quality of courses. A negative coefficient means that ethnic minority students are lower matched than White students, choosing courses which are lower quality, for a given level of attainment. A positive coefficient means that ethnic minorities are higher matched than White students, choosing courses which are higher quality, for a given level of attainment. For example, a coefficient of -5 means that ethnic minority students attend courses which are 5 percentiles lower quality than White students from the same attainment quintile. A coefficient of +5 means that ethnic minority students attend courses which are 5 percentiles higher quality than White students from the same attainment quintile.

Ethnic gaps in match quality – points-based (Figure 7a)

- Finally, we present gaps in match for White students versus students from ethnic minority backgrounds. For simplicity, we present results only after conditioning for characteristics and prior attainment.
- Figure 7a suggests that, among low attainers, there are no apparent differences in extent of points-based overmatch. Among high attainers, however, we can see that, with the exception of Black Caribbean and Black students from other backgrounds, ethnic minority students are higher matched than White students. Given that the baseline match index is around -45 percentiles for the highest attaining quintile of students (i.e. high-attaining students typically undermatch by 45 percentiles), this finding means that high-attaining ethnic minority students are still undermatched, but by around 5 percentiles less than White students.

Figure 7b: Ethnic gaps (ethnic minority-white) in match earnings-based



Notes on how to read these charts: the points on the chart show the ethnic gap in match (ethnic minority-White) in percentiles. A coefficient of zero would mean there is no ethnic gap, so that both ethnic minority and White students choose the same quality of courses. A negative coefficient means that ethnic minority students are lower matched than White students, choosing courses which are lower quality, for a given level of attainment. A positive coefficient means that ethnic minorities are higher matched than White students, choosing courses which are higher quality, for a given level of attainment. For example, a coefficient of -5 means that ethnic minority students attend courses which are 5 percentiles lower quality than White students from the same attainment quintile. A coefficient of +5 means that ethnic minority students attend courses which are 5 percentiles higher quality than White students from the same attainment quintile.

Ethnic gaps in match quality – earnings-based (Figure 7b)

- Interestingly, ethnic gaps in earnings match are far more extreme than in points-based match, with ethnic minority students choosing higher earning courses than White students, in general.
- This is true for both high and low attainers, though the ethnic minority advantage in earnings match seems larger for high attainers.
- Asians, Black Africans, Chinese, mixed and students from other ethnic backgrounds all undermatch less than Whites – with the gaps around three times larger than in points-based match.
- This evidence suggests that White students select onto courses with lower average earnings than ethnic minority students.

8 Examining the severity of mismatch

So far, we have seen that there is a significant degree of mismatch in the UK higher education system, and that there are stark inequalities in how well students match themselves to courses. Students from disadvantaged backgrounds undermatch more and overmatch less than their advantaged counterparts. And female students undermatch more than males when we measure course quality according to earnings.

In this section, we aim to understand whether these inequalities hold among the extremely undermatched, and the extremely overmatched. To be clear on who we are focusing on in this exercise, Figures 8a and 8b show the distribution of mismatch for low and high attainers, by SES and gender. As we can see from these figures, low attainers are skewed to the right of zero – being overmatched – and high attainers skew to the right – being overmatched.

For the purposes of this exercise, we examine SES and gender gaps in match across this match distribution, with particular interest in the extremes (the 10th and 90th percentiles of match), as highlighted in the figures. Hence, for low attainers, who overmatch, we are interested in the least overmatched (around -20 to 0 match index on average), and the most overmatched (around 80-100 match index on average). For high attainers, who undermatch, we are interested in the most undermatched (around -100 to -80 match index on average), and the least undermatched (around 0 to 20 match index on average).

Figure 8a: Mismatch distribution, by SES and attainment

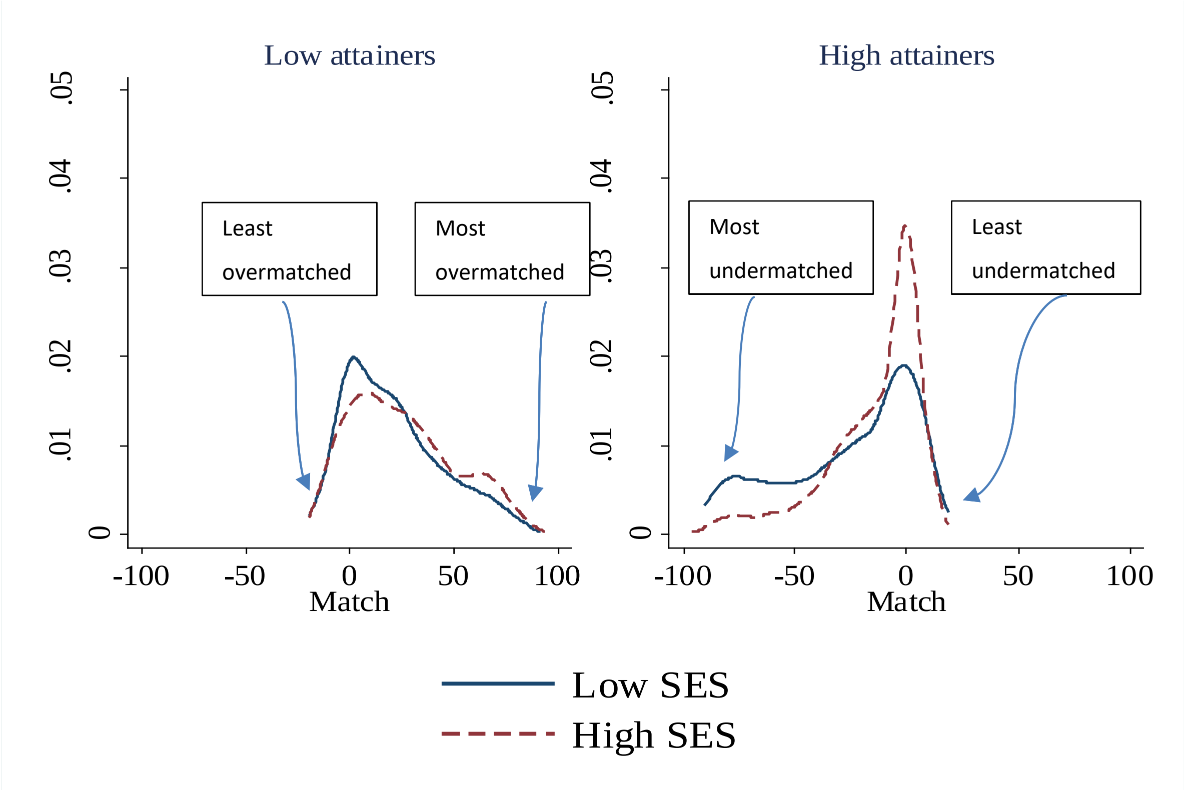
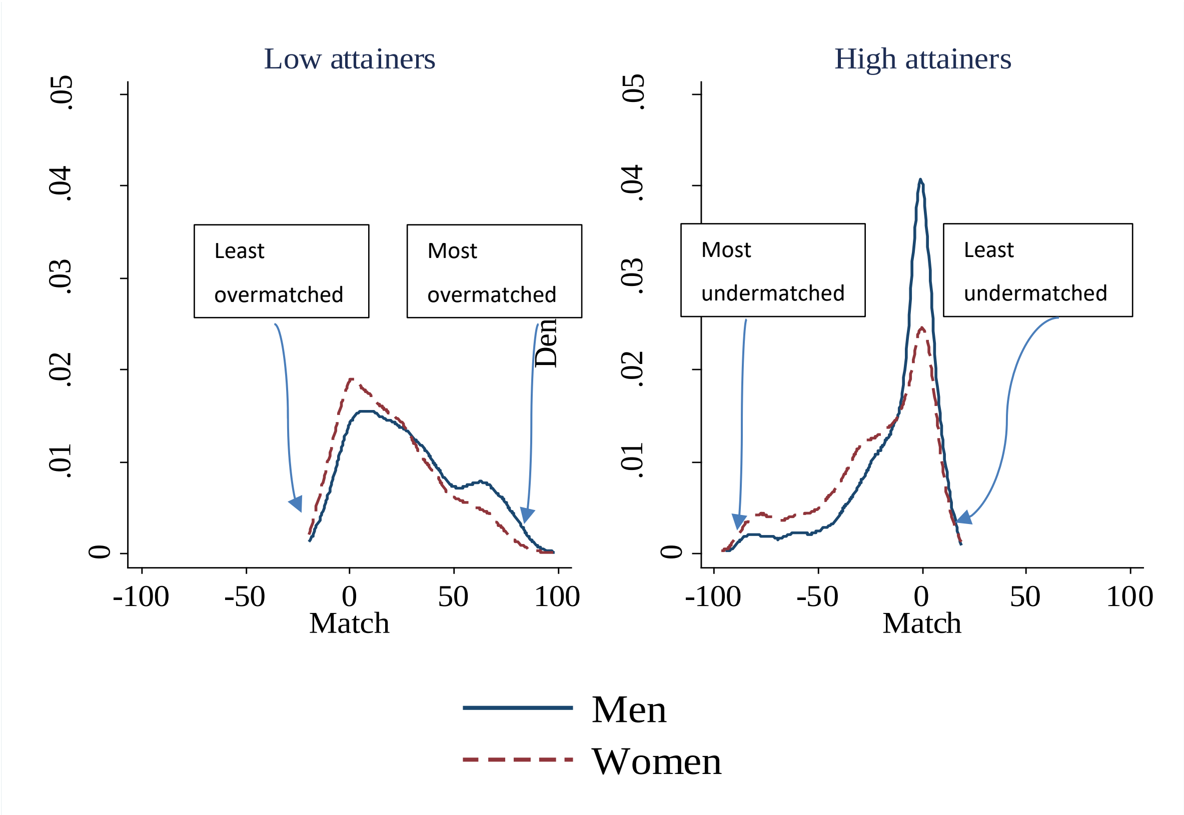


Figure 8b: Mismatch distribution, by gender and attainment



Having established the groups of students we are interested in, in Figures 9a and 9b we present coefficients from unconditional quartile regression (see Appendix Figure 32). These figures essentially express the SES gaps (9a) and gender gaps (9b) in match, across the match distribution seen above in Figures 8a and 8b. For low attainers, we are spanning from the least to the most overmatched students, and for high attainers spanning from the most to the least undermatched.

These charts reveal the following insights:

SES gaps in match severity (Figure 9a)

- Starting at the far left of the first panel of Figure 9, we can see SES gaps in match quality for the least overmatched. Here, there are clearly only small differences in overmatch by SES. However, as we move up the distribution of match from the least overmatched to the most overmatched, these differences become more pronounced. By the time we reach the most overmatched – i.e. the students who are attending courses that are above them in terms of quality by as much as 80 percentiles in the distribution, the gaps are stark. Here, low SES students overmatch by 10 percentiles less than high SES pupils.
- Thus, these graphs show that even within the group of low achieving students who manage to significantly overreach themselves in terms of the course they eventually access, students from richer backgrounds still manage to reach further than poorer students.
- For high attainers (the second panel in 9a) we turn our focus to the extremes of undermatch. Here, the picture is even starker among the most severely undermatched. For the most undermatched students (at the far left of the chart) – i.e. those who attend courses substantially lower in quality than they could have, low SES students undermatch by 30 percentiles more than high SES students.
- Appendix Figure 2 shows that this pattern holds for our points-based match measure, though for the low-attaining students the SES gaps are generally smaller and sometimes go to zero.

Gender gaps in match severity (Figure 9b)

- Repeating the exercise for gender, rather than SES gaps tells a very similar story. Starting with the least overmatched, we see only small gender differences. As we move towards the most overmatched – those attending courses of substantially higher quality than their academic credentials would predict – we see that males reach significantly higher quality courses than females – with the male female gap reaching almost 15 percentiles.

- For high attainers (the second panel in 9b) at the extremes of undermatch, again we see stark gender gaps. Among the most severely undermatched the gender gap reaches 20 percentiles.
- Appendix Figure 3 shows that this pattern holds for our points-based match measure for high attainers, though, for low attainers, the SES gaps are insignificant across the match distribution.

Figure 9a: Severity of match, by attainment and SES

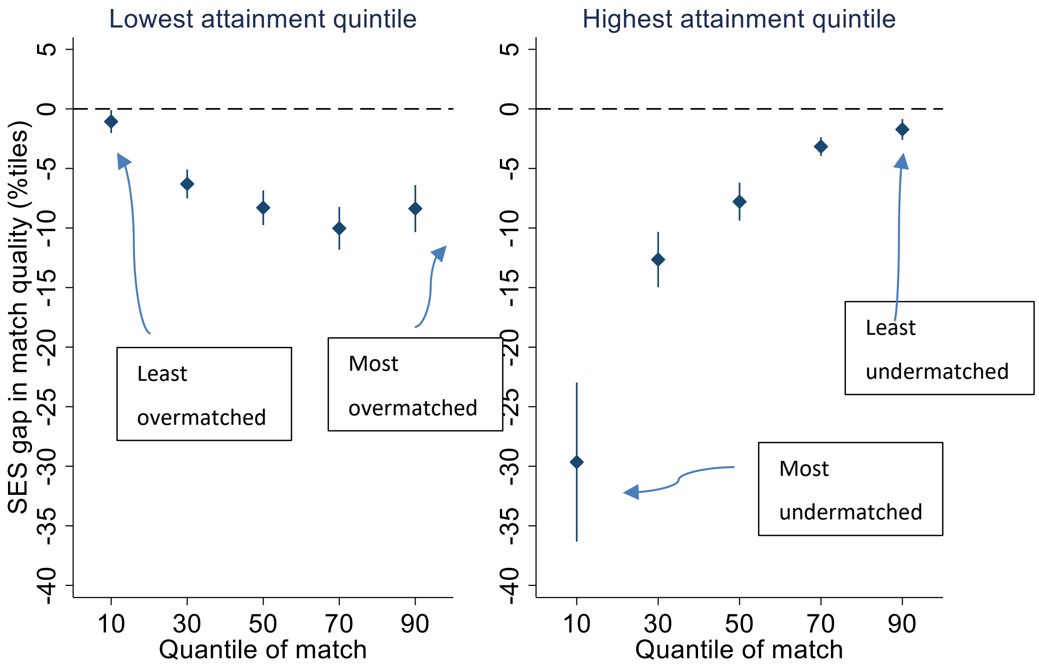
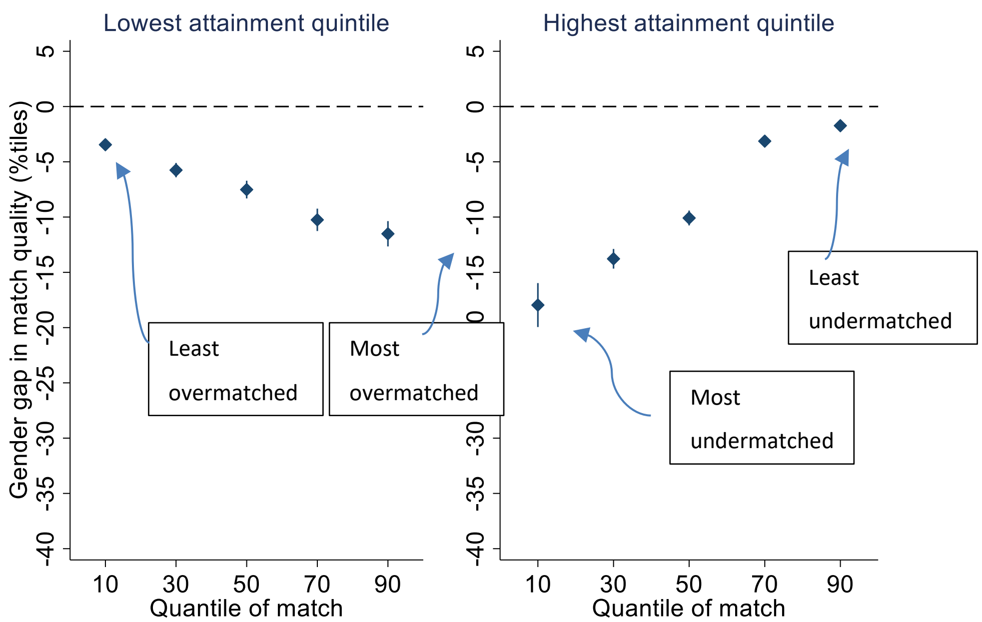


Figure 9b: Severity of match, by attainment and gender



9 What drives mismatch?

We now turn our attention to possible explanations for these SES and gender inequalities in earnings match. Driven by literature in the area (Dillon and Smith, 2017; Hoxby and Avery, 2012), we explore three possible factors which may potentially drive students to choose a lower quality course than they could have – subject choice, geographical factors, and school factors.

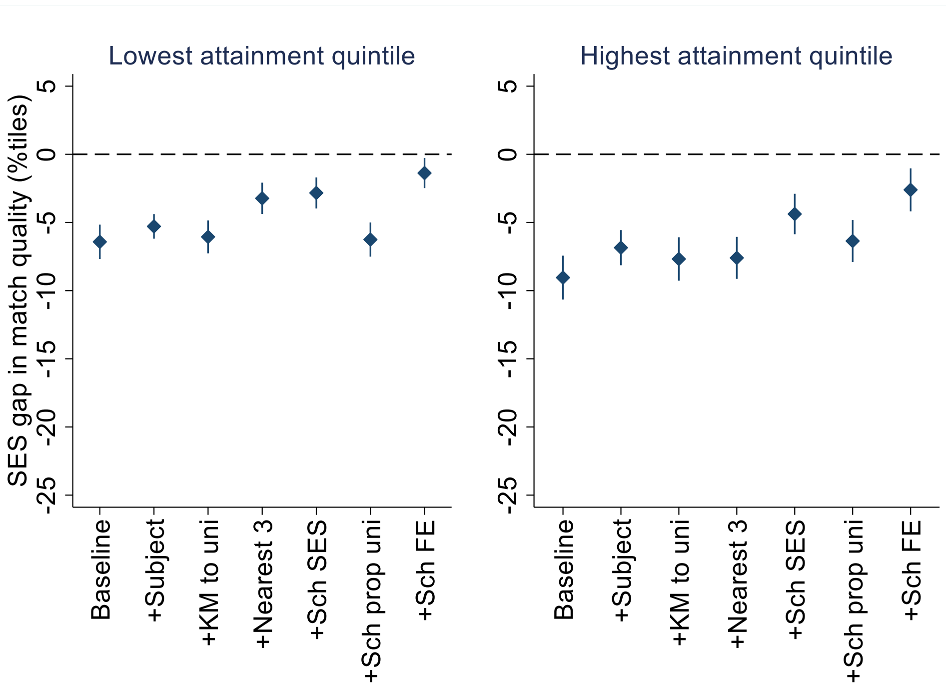
Table 5: potential drivers of mismatch

Subject choice	Geography	School
Degree subject of study	Km to nearest university	Socio-economic mix of the school
	Nearest 3 universities to home	Proportion of pupils going to university
		School attended

To achieve this, we run a series of multivariate regressions, with match as our outcome variable, conditioning on a range of measures to investigate whether our SES and gender differences can be explained by the inclusion of these variables, shown in Figures 10 and 11.

In these charts each individual factor is shown relative to the baseline SES and gender gap, as seen in Figures 5 and 6 (which takes account of demographics and prior attainment).

Figure 10: SES gaps in earnings match: drivers



Notes on how to read these charts: the points on the chart show the socio-economic gap in match (low SES-high SES) in percentiles taking account of the controls listed on the x-axis (with baseline being the coefficient reported in figure 5b). A coefficient of zero would mean there is no SES gap, so that both low and high SES students choose the same quality of courses. A negative coefficient means that low SES students are lower matched than high SES students, choosing courses which are lower quality, for a given level of attainment. For example, a coefficient of -5 means that low SES students attend courses which are 5 percentiles lower quality than high SES students from the same attainment quintile, after taking account of the controls listed on the x-axis. If the coefficient is smaller than at baseline, this implies the control listed on the x-axis is a potential driver of the SES gap.

9.1 Drivers of the SES gaps in match

Subject studied

- We first examine the potential impact of subject studied at university. By conditioning on subject studied, we are exploring whether low SES students are choosing subjects that command lower earnings than high SES students, or whether they are attending lower quality institutions, regardless of subject studied.
- Figure 10 illustrates that very little of the SES inequalities in match are driven by the subjects that people study at university. Most of these gaps remain intact, indicating that even when students are of similar attainment and studying the same subject, low SES students study it at a lower earning institution. Thus, we can conclude that a key driver of SES inequalities in match is the choice of institution attended.

- The picture is very similar when using our attainment based measure of match (Appendix Figure 4).

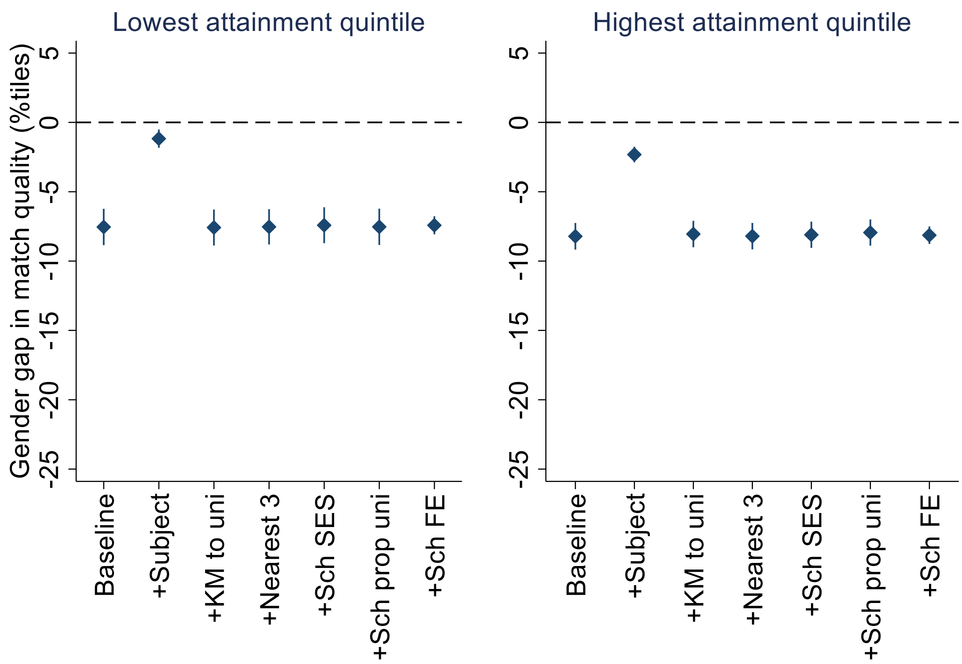
Geographical factors

- Previous studies of match in the US have highlighted the importance of proximity to colleges as a key driver of match (Hoxby & Avery, 2012). We therefore explore two alternative specifications of proximity to university: distance to the university attended and each of the nearest three universities to home. Figure 10 indicates that the nearest three universities to home can account for some of the SES gap for low attainers, but neither measure of distance to university can account for the SES gap in match for high attainers
- Again we see a similar picture for attainment match (Appendix Figure 4).

School level factors

- We go on to consider three alternative measures of school-level possible drivers: the SES mix of the school attended, the proportion of the school attending university and a black-box school fixed effects model. Figure 10 shows that the proportion of the school attending university does little to account for SES gaps in match for low and high-attaining pupils.
- This is in contrast to the US literature which highlights the importance of role models in the form of previous cohorts attending college (Dillon & Smith, 2017, Black et. al. 2015).
- The socio-economic demographic mix of the school can account for around half of the SES gap in match for both high and low attainers, while controlling for the school attended (school fixed effects) almost completely eradicates the SES gap. This suggests that other unobserved school level factors, for example advice offered on subject choices, guidance on institutions, school sorting, and location, account for much of the difference in match quality between high and low SES students.

Figure 11: Gender gaps in earnings match: drivers



Notes on how to read these charts: the points on the chart show the socio-economic gap in match (low SES-high SES) in percentiles taking account of the controls listed on the x-axis (with baseline being the coefficient reported in figure 5b). A coefficient of zero would mean there is no SES gap, so that both low and high SES students choose the same quality of courses. A negative coefficient means that low SES students are lower matched than high SES students, choosing courses which are lower quality, for a given level of attainment. For example, a coefficient of -5 means that low SES students attend courses which are 5 percentiles lower quality than high SES students from the same attainment quintile, after taking account of the controls listed on the x-axis. If the coefficient is smaller than at baseline, this implies the control listed on the x-axis is a potential driver of the SES gap.

9.2 Drivers of the gender gaps in match

Subject studied

- When we look at Figure 11, it is clear that most of the gender gap can be accounted for by the subject studied at university. For low attainers, the gender gap is almost completely eradicated when estimated within subject grouping, suggesting that the reason lower attaining women are found at lower earning courses is because they choose subjects that command lower labour market rewards. However, even after controlling for subject of study, high-attaining women still undermatch to a greater extent than high-attaining men, choosing lower ranked institutions in terms of earnings.
- There is very little gender gap to be accounted for using our points-based measure of match, yet Appendix Figure 5 shows clearly that none of our potential mediators explain the remaining small gap in undermatch for high-attaining women.

Geographical factors

- We see no significant influence of geographical factors in driving the male female gap in match.

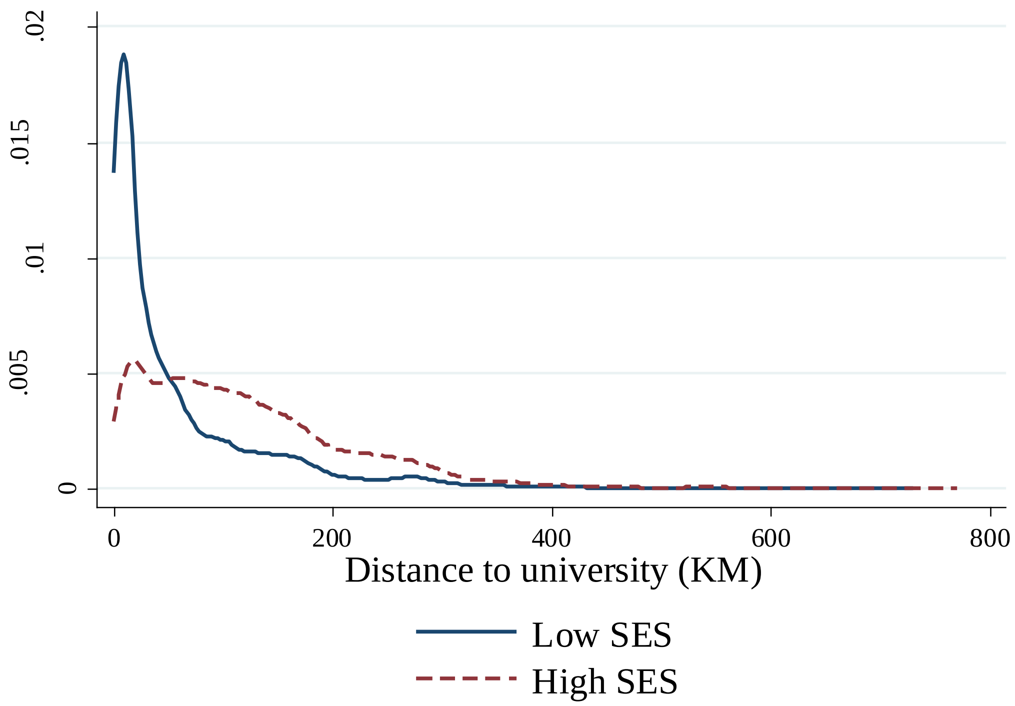
School level factors

- Finally, none of our school level factors appear to be potential mechanisms in the gender gap in earnings match. This is perhaps unsurprising, since males and females are not typically segregated geographically, or across schools.

Distance to university

As discussed above, we found no influence of any of our geographical factors in explaining the SES or gender gap. However, as Figure 12 shows, there is a substantial SES gap in distance to university attended. In particular, low SES students are far more likely to be found at universities close to their home location[15]. This chart shows students of all levels of attainment, but the results are similar if just examining high-attaining students.

Figure 12: Distribution of distance travelled to university for low and high SES students



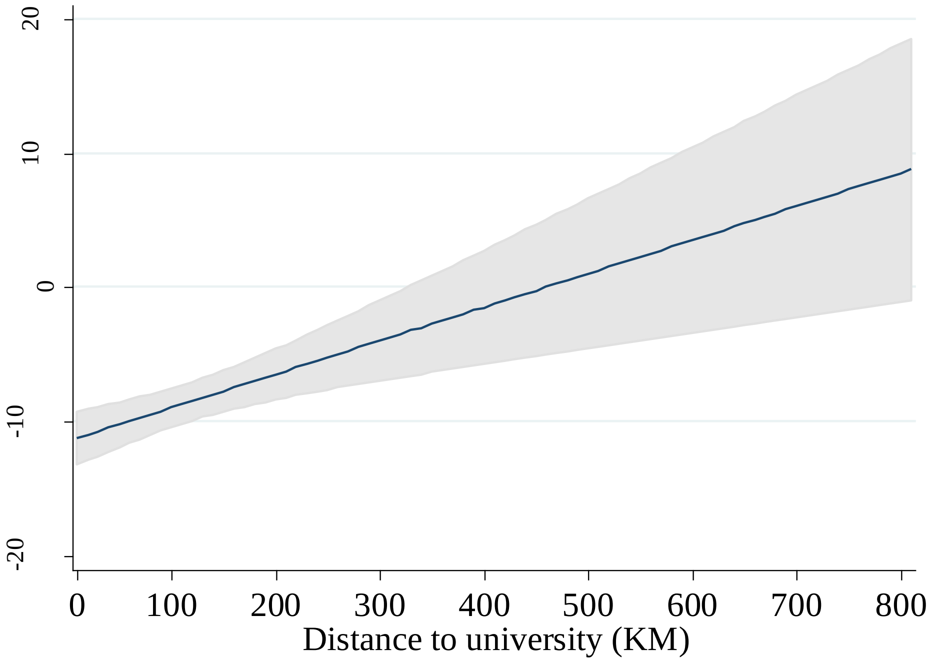
[15] We define their home location using the centre of the home Lower Super Output Area level, in the absence of information on their home postcode

To explore the role of distance travelled to university on match further, Figures 13 (and Appendix Figure 6) shows SES gaps in match, according to distance travelled (after our usual set of controls) for high-attaining students.

This shows that the SES gaps in match actually vary by distance travelled. For those who attend university near home, low SES students undermatch more, attending lower ranked universities in or near their hometowns, such as Liverpool John Moores, Manchester Met and University of Teeside. High SES students on the other hand undermatch less, with those attending unis near home typically studying at University of Manchester, University of Liverpool or University of Newcastle. This could be driven by information constraints, with low SES students being less likely to be informed about the differences between the universities. Alternatively, this could be driven by preferences, due to the perceived intake of each institution, and how well the student thinks they will fit in. It is important to remember here that all of these students are in the top 20% of (difficulty adjusted) achievement at age 18 and so we are comparing students with similar prior attainment, that choose to stay close to home, but make quite different choices about the institution they attend based on SES.

In contrast, low SES students who are willing to travel to a university further away from their home location are actually as equally well matched as high SES students. A possible explanation is that these students could be better informed about university quality, and are therefore more motivated to move away from home.

Figure 13 Conditional SES gap in earnings-based match by distance to university attended (high achieving students only)



We now examine the role of other preferences and attitudes that students may have – including forward planning, preferences around prestige, and university choice behaviour – on match. For this exercise, we turn to linked NPD-Next Steps data, as briefly described in Section 4.

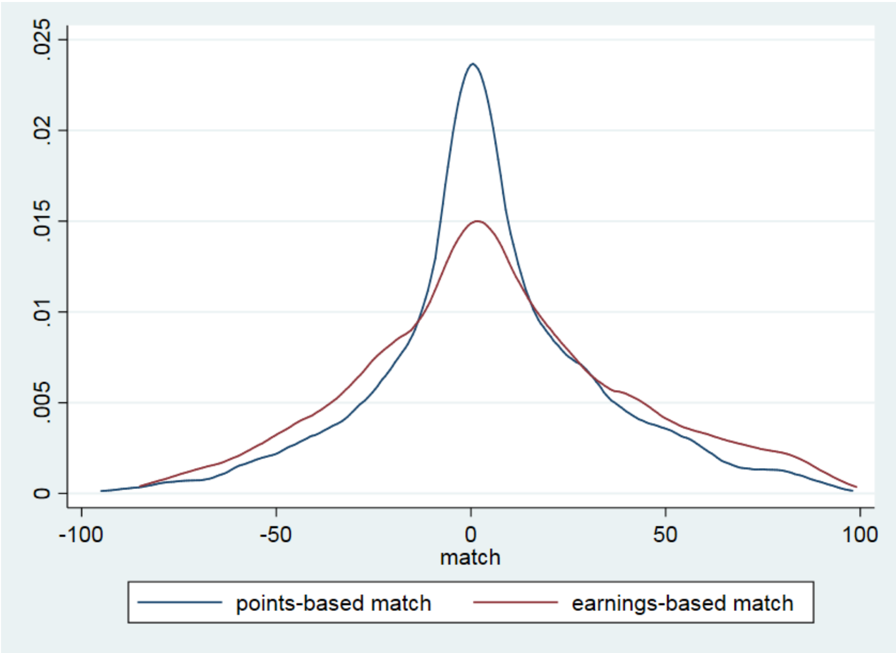
Preferences and attitudes of mismatched students

As described in Section 4, in order to understand the aspirations and preferences of mismatched students, we use linked Next Steps-NPD data on students from the same cohort.

While Next Steps contains a rich set of covariates, sample sizes are obviously more limited than in our administrative data. This is problematic for calculating our course quality measures since many courses only have a few students studying in them. We therefore use the individual and course position in the *national* distribution of achievement (and earnings) from our population data to calculate match.

Figure 14 shows the distribution of our points-based and earnings-based match using the NPD-Next Steps sample.

Figure 14 Points-based and earnings-based measures of student-course match



Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

These distributions are similar to that found in our main NPD-HESA dataset, with our earnings-based measure of match rather more spread out than our points-based measure. Table 6 shows our Dillon and Smith measure of match (i.e. students undermatched if they are 20 or more percentiles below the position of their course, and overmatched if they are 20 or more percentiles above the position of their course) for both measures of match.

Table 6: Points-based and earnings-based measures of student-course match

	Undermatched (%)	Matched (%)	Overmatched (%)
Points-based	16	59	25
Earnings-based	22	47	31

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

This indicates that both our measures of match show similar proportions of undermatched students compared to our original NPD-HESA sample (the latter containing 15% undermatch for points-based, and 25% undermatch for earnings-based). However our Next Steps-NPD sample over-represents the extent of overmatch, both for earnings-based and points-based match (where our NPD-HESA sample showed 15% overmatch for points-based and 23% for earnings-based). This is likely due to the small sample, and possible selection effects.

Given these concerns, and the fact that our nationally representative sample indicates that, after controls, there is only a very small gaps in overmatch, we focus for the rest of this section on undermatched students – i.e. those from the top 20% of the attainment distribution.

Table 7: Points-based and earnings-based measures of student-course match – high performing students

	Undermatched (%)	Matched (%)
Points-based		
Male	20	80
Female	21	79
Low SES	35	64
High SES	17	83
Earnings-based		
Male	29	71
Female	37	63
Low SES	47	53
High SES	31	69

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Table 7 highlights the SES and gender gaps in undermatch for high attainers in NPD-Next Steps. The patterns are very similar to those in the NPD-HESA administrative sample. For points-based match, we see no evidence of gender gaps, but evidence that low SES students are much more likely to undermatch than high SES students, even among those from the highest quintile of attainment.

For earnings-based match, we see large gender gaps, with high-attaining women much more likely to undermatch on earnings than high-attaining men just as we found in our administrative data. And finally, we find that high-attaining low SES students are more likely to be undermatched than high-attaining high SES students as in our points measure, for earnings-based match.

These similarities are very reassuring, giving us confidence that despite the limited sample sizes in the NPD-Next Steps data, we are capturing similar trends. With this in mind, we therefore go on to examine the attitudes and preferences of high-attaining undermatched students.

We use multiple regression analysis for this exercise, with our mismatch index as our outcome variable, conditioning on a range of measures related to attitudes, preferences and advice and guidance (entered in groups, with each group as a separate regression). In each model we also condition on our SES and gender gaps.

Tables 8 and 9 present these regressions for both our points and earnings-based measures of match. In each case, since we run these regressions only for high achieving students, the regression coefficients can be interpreted as the impact on going from undermatch to match.

Table 8: Regression analysis – impact of attitudes and preferences on points-based match - high-attaining students

	(1)	(2)	(3)	(4)	(5)
VARIABLES	raw	believe apply/get in	prestige/ location	whether first choice	whether talked to teachers
low SES	-11.35***	-10.80***	-9.938***	-11.55***	-10.99***
	(2.686)	(2.640)	(2.652)	(2.720)	(2.701)
medium SES	-4.300**	-3.032	-4.327**	-4.413**	-4.189**
	(2.115)	(2.072)	(2.074)	(2.119)	(2.115)
gender	0.834	0.660	1.323	0.858	0.960
	(1.538)	(1.503)	(1.530)	(1.543)	(1.539)
think will apply		12.27***			
		(2.363)			
think will get in		1.616			
		(1.671)			
importance: location			-4.703**		
			(1.889)		
importance: good uni			7.255***		
			(1.558)		
importance: good sub			-2.340		
			(2.306)		
uni not first choice				4.047	
				(2.756)	
sub not first choice				-5.923	
				(5.002)	
talked w teachers					2.483
					(1.580)
Observations	642	642	642	642	642
R-squared	0.031	0.085	0.079	0.037	0.035

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

First examining Table 8 – points-based match – column 1 highlights the significant SES gap in points-based match, and also that there is no evidence of a gender gap in this measure of match.

Next, in column 2, we look at correlations between match and the students beliefs – taken at age 15/16 – about whether they will apply to university, and whether they will actually get in. It is interesting that there is a significant positive relationship between whether a student thinks they will apply and how well matched they are to their future course – but no such relationship between students confidence of their future success and match. On average, students who answer that they are 'very likely' to apply to university are 12 percentiles better matched than students who report to be 'fairly likely' (or anything below this) to apply to university. This may be evidence of future planning among students who believe they will apply to university at age 15/16. Note that these mediators do not particularly drive either our SES or gender gaps but independently predict differences in match quality.

We next go on to look at students university preferences, examining the perceived importance of location, university prestige, and subject prestige – measured at age 17/18 when the student is undertaking A levels or post-compulsory qualifications. Here, we find an important role for location preferences – students who stated that location of the university was important (regardless of whether the university is near or far from home) were lower matched by around 4 percentiles. The implication is that students who are at least in part influenced by the location of the university may pay less attention to the academic fit. This complements our analysis on the importance of distance travelled in the previous section, again highlighting the penalty to being constrained or influenced by location.

On the other hand, students for whom university prestige was important are found to be higher matched by some 7 percentage points – though there is no such impact for their beliefs about subject prestige. It is again interesting to note that these mechanisms do not influence the SES or gender gaps. In the case of our location measure this is particularly notable, since this reaffirms our finding that geography does not reduce SES gaps in match directly, but has a more complex interaction with SES and match.

Our third set of regressions examine whether the university and subject the student is studying (at age 18/19) was not their first choice (due to not getting an offer, rather than low grades). If students who did not get their first choice are lower matched (more undermatched), the implication would be that university acceptances are impacting match, rather than that students are choosing less elite courses. As we can see, there is no significant relationship between choice and match, implying that the issue is not likely to be driven by universities turning students down, but rather more likely to be driven by the fact that students are choosing undermatched universities. This is in keeping with US findings (e.g. Smith et al, 2013).

Finally we look at whether there is any influence of teacher advice and guidance on match, and we find no difference in match between those who talked to teachers and those who did not. Again, like all the other mechanisms, SES and gender gaps are not affected.

Table 9: Regression analysis – impact of attitudes and preferences on earnings match - high-attaining students

	(1)	(2)	(3)	(4)	(5)
VARIABLES	raw	apply/get in	prestige/ location	whether first choice	whether talked to teachers
Low SES	-10.2***	-9.339***	-8.916***	-9.670***	-10.03***
	(3.032)	(3.009)	(3.035)	(3.060)	(3.055)
Medium SES	-2.287	-1.225	-2.274	-2.367	-2.189
	(2.387)	(2.361)	(2.374)	(2.383)	(2.392)
gender	-2.907*	-2.907*	-2.462	-2.781	-2.842
	(1.736)	(1.713)	(1.751)	(1.735)	(1.740)
think will apply		10.28***			
		(2.693)			
think will get in		2.878			
		(1.904)			
importance: location			-4.622**		
			(2.163)		
importance: good uni			4.653***		
			(1.783)		
importance: good sub			-2.607		
			(2.640)		
uni not first choice				3.964	
				(3.100)	
subject not first choice				-15.38***	
				(5.626)	
talked w teachers					1.779
					(1.786)
Observations	642	642	642	642	642
R-squared	0.022	0.060	0.045	0.035	0.024

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Source: University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

In Table 9 we repeat this exercise for our earnings-based measure of match. Again our raw regression in column 1 highlights the significant SES and gender gaps in earnings-based match, with low SES students and women choosing lower returns courses.

Looking at the other sets of controls as above, in general we find very similar results. We again see a significant positive relationship between whether a student thinks they will apply and how well matched they are to their future course – but no importance of future success. We also observe the same patterns for subject and university preferences, with location and prestige of university both significantly correlated with earnings-based match, in a similar manner as for points-based match.

Looking next at whether university and subject were the students' first choice, we do find a role for subject in relation to earnings mismatch. In particular we find that students who did not get their first choice of subject due to not getting an offer (rather than low grades) were significantly more undermatched. The implication is that students who are rejected from their first choice subject (but not institution) are more undermatched in terms of earnings, but not points. Finally, we again find no role for teacher advice and guidance.

In summary, we find some evidence that students who are undermatched have different attitudes and preferences towards university, and whether these attitudes may influence their eventual match. We see that students who are forward planning at age 15/16 are better matched. We also find that those who have some idea of university prestige and its importance are better matched, while those who may be influenced by the location of a university are more undermatched. Finally there is some evidence that students who do not get their first choice subject offer are more undermatched in terms of earnings match, but not points match. This implies that students who are thinking ahead, and perhaps researching factors related to university quality, are likely to be better matched, and that any kind of policy intervention should be aimed here.

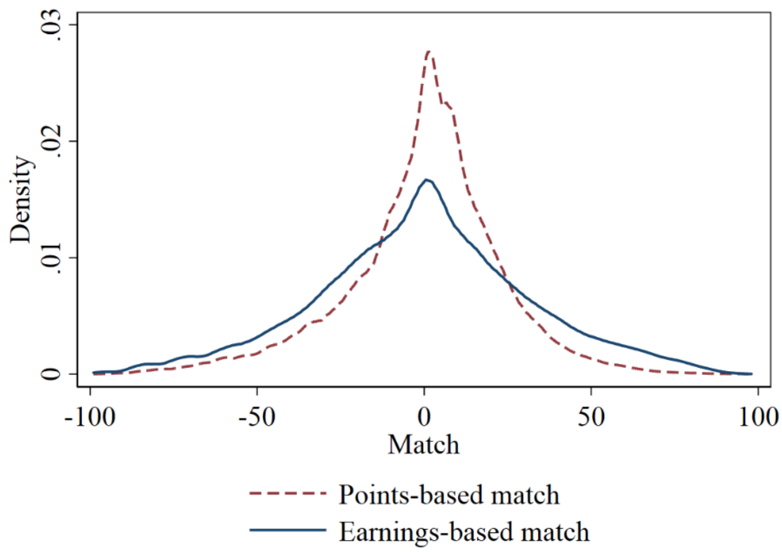
10 University outcomes of mismatch

In this section we detail the association between mismatch and outcomes during and at the end of university. As detailed in the introduction, while there has been a burgeoning literature on the returns to individual qualifications (McIntosh, 2006, Crawford, 2014), and the returns to different university subjects and institutions (Belfield et al., 2018, Hussain et al., 2009, Chevalier, 2011, Walker and Zhu, 2013), less is understood about the consequences of how well matched students are to their university course. The likely outcomes from mismatch are ambiguous. Overmatched students may benefit from being surrounded by higher quality peers on higher quality courses. Yet they may also struggle from being ill-prepared for the level of knowledge required on higher-quality courses. Similarly, undermatched students may benefit from being the big fish in a small pond. But they may also be held back by lower quality peers, lower quality teaching and lower expenditure per head.

Dillon and Smith (2019) consider the impact of match on degree completion and later labour market earnings, using survey data from the US, and find some evidence of positive effects of matching on degree completion and later earnings. Nothing is known beyond this one study about the potential consequences of mismatch.

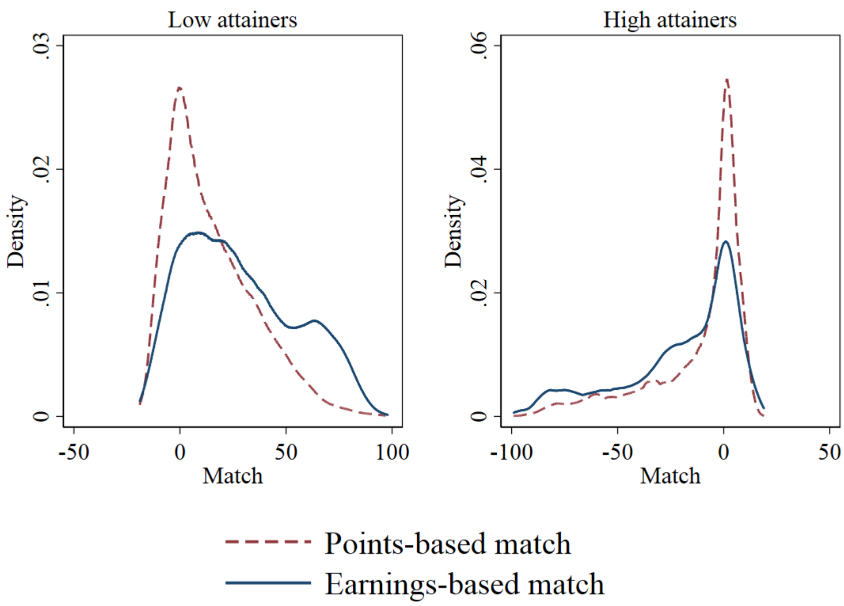
Mismatch is defined in the same way here as it is in the previous sections. We focus on the same cohort of students, although our sample of data is slightly different as we now require linked outcome information to analyse the association between mismatch and outcomes. As there are some slight differences in the sample, we replicate Figure 2 for this sample of data (see Figure 15 below). Reassuringly there are very similar proportions of students who mismatch in the outcomes sample with 15% of our points-based measure and 24% of our earnings-based measure under and overmatched.

Figure 15: Points-based and earnings-based measures of student-course match



Given what we have learned about the differences in match across the distribution of attainment, we continue our focus here on low attainers and high attainers. We also move from using a continuous measure of match to focusing on a binary definition as defined in Table 3 (20 percentiles +/- zero). This is useful in this setting as it simplifies our analysis: for low attainers we can compare those who overmatch to those who match and for high attainers we can compare those who undermatch to those who match. The comparison group is therefore constant across our analysis but this approach allows us to separate mismatch into these two distinct types. As motivated above, it is reasonable to assume that the outcomes for undermatched students compared to matched students might be quite different to the outcomes for overmatched students, relative to matched students.

Figure 16: Points-based and earnings-based measures for low and high attainment



As can be seen from Figure 16, the proportion of low-attaining students who overmatch is higher than in the overall sample, as would be expected, with 34% overmatching for our points-based measure and 55% overmatching for our earnings-based measure. The second peak in overmatched students for the earnings-based measure of match represents courses that have a high return but lower entry requirements, such as business, engineering and technology related fields, and nursing. For high-attaining students a similar pattern emerges for undermatched students, with 24% undermatched on our points-based measure and 39% undermatched on our earnings-based measure.

It is important to remember, when comparing outcomes across our different measures of match, that these alternative measures create quite distinct groups of students who are mismatched. While the previous sections have highlighted the difference in characteristics between those who over- and undermatch *compared to those who match* for both our points-based and earnings-based measures, Table 10 illustrates the main differences, in terms of characteristics, subject studied and institution attended, *between* those who over- and undermatch *across* our measures of match.

Table 10: Differences in the characteristics of those who mismatch across measures of points-based and earnings-based match

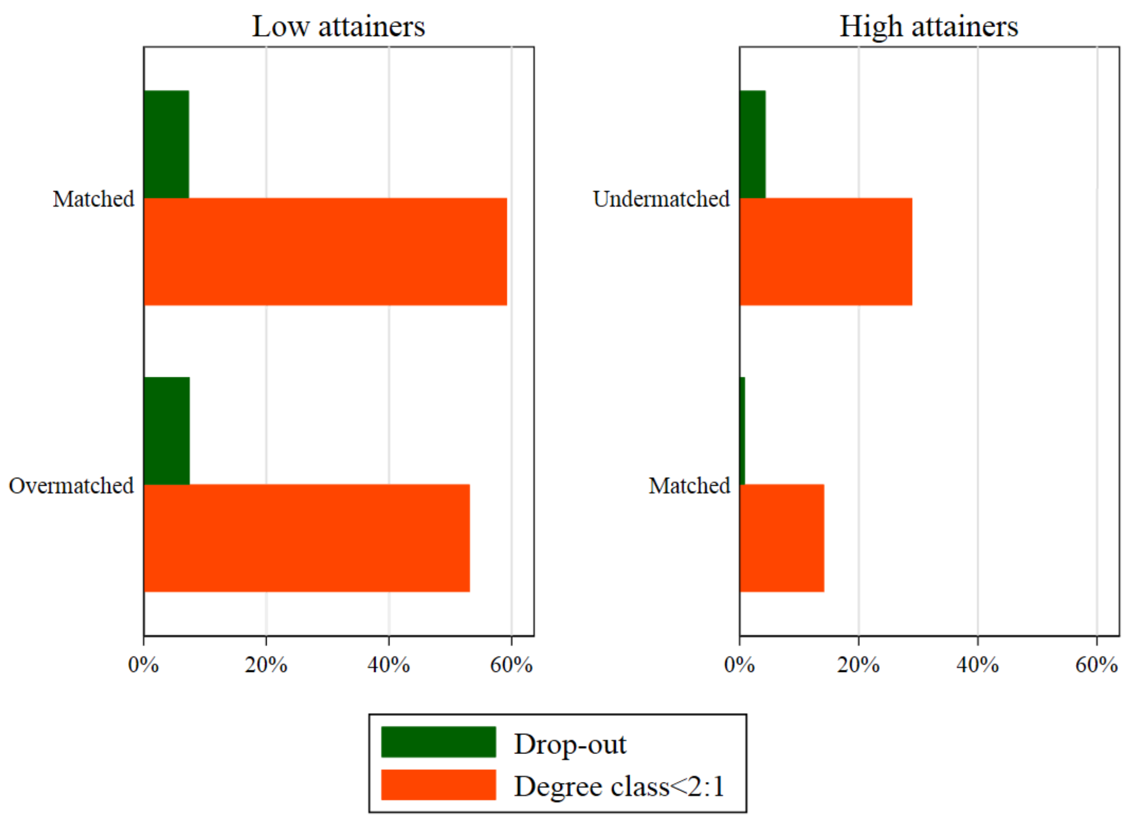
	Overmatched low attainers		*Undermatched high attainers*	
	Points-based match	**Earnings-based match**	**Points-based match**	**Earnings-based match**
Female	51	45	63	66
High SES	70	68	69	75
Low SES	30	32	31	25
Ethnic minority	17	27	15	10
STEM	38	55	40	41
Social Science	29	38	29	20
Arts and Hum.	31	6	28	37
Post-92	76	83	76	47
Old non-RG	20	13	20	27
Russell Group	5	4	4	26
A lev. percentile	10	10	90	90
Total	34	45	24	39

Overmatched students, in terms of achievement-based measures of match, are more likely to be White students, who are broadly spread across subjects but are more likely to study arts and humanities, at older non-Russell Group institutions, relative to overmatched students using earnings-based match. Earnings-based overmatched students in contrast are a higher proportion of ethnic minority students, who predominantly study STEM (and to a lesser extent social science), and at a post-92 institution.

Undermatched students, in terms of our points-based measure of match, are slightly more likely to be low SES, ethnic minority students, who study a range of degrees but are more likely to study social science, at predominantly post-92 institutions relative to students who undermatch in terms of our earnings-based measure of match. In contrast, those who undermatch on our earnings measure of match are from slightly higher SES, White families, and are more likely to study arts and humanities subjects at Russell Group institutions.

We begin by presenting some basic graphs to highlight the difference in university outcomes between matched and mismatched students, before moving to attempt to control for observable differences between matched and mismatched students using a multivariate regression approach. In this setting, we control for different groups of observable characteristics in stages, beginning with demographic differences that we saw were important drivers of mismatch in the previous section, such as parental background, gender and ethnicity. We then control for the school that pupils attended – effectively comparing individuals who match and mismatch from similar backgrounds from the same school. In the fourth column we control for the broad university group category that the individual attends. This allows us to compare individuals who attend higher quality universities (such as Russell Group universities) with similar individuals who attend similar quality (but not the same) institution.[16] Finally we consider the potential role of subject choice in our findings by looking at differences in outcomes between matched and mismatched students by the choice of course, considering whether the student studies STEM, social science, or arts and humanities.

Figure 17: University outcomes by match status for low and high attainers (points-based match)



[16] The groups are Post-1992, older non-Russell Group institutions, and Russell Group universities

We consider two outcomes at university, whether the students drops out of university at any point during the first 3 years of study, and whether the students is graded lower that a 2:1 on degree completion (a 'lower class degree'). 7% of low achievers drop out of university and 56% achieve less than a 2:1. For high attainers, 2% drop out of university and 17% receive a grade below a 2:1 upon completion. Figure 17 shows that:

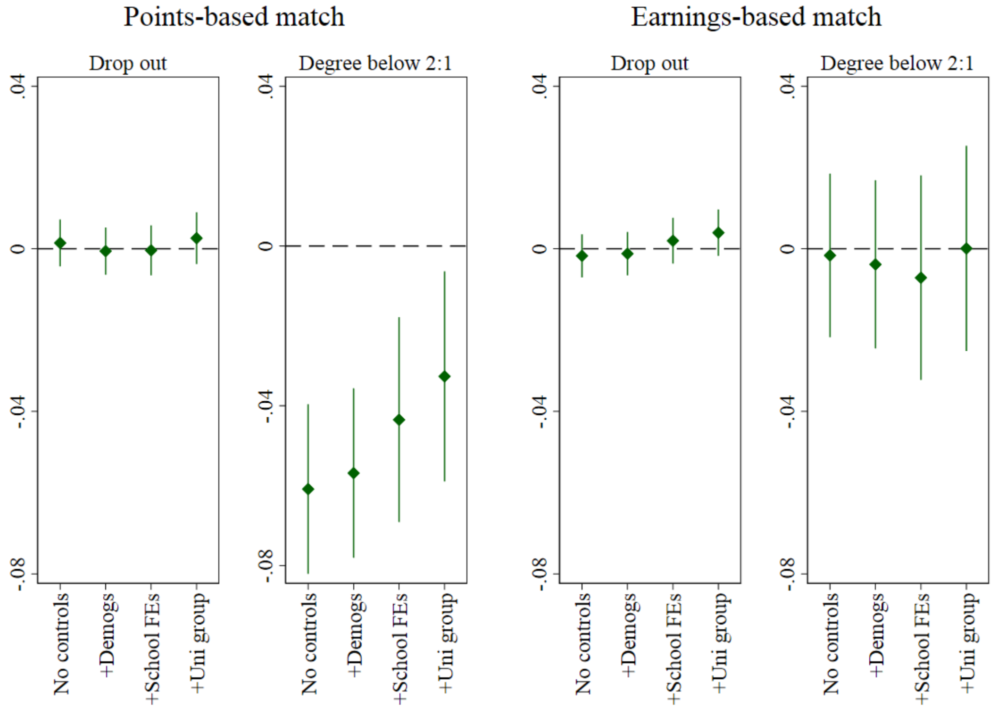
- Low-attaining students have a similar chance of dropping out whether they were matched or overmatched to their course.
- There is a difference in the proportion achieving a lower class degree with low-attaining overmatched students slightly less likely to receive a grade below a 2:1 compared to matched students (52% of overmatched students relative to 58% of matched students).
- There are more marked differences for high-attaining students, with matched students being less likely to drop out (1% vs 4%) and receive a lower class degree (13% vs 29%), relative to undermatched students.

Figure 18 and 19 present results from a regression approach, where we can control for other observable characteristics that might explain the raw differences in outcomes between those who are well-matched to their course relative to those who are mismatched. For completeness, we present results from our points-based match measure in the first two boxes, and the results from our earnings-based match measure in the second two boxes, highlighting any differences in outcomes across the two alternative match measures as they arise. As before the dots indicate the estimated gap in outcomes while the bars indicate the confidence intervals around this. Note that confidence intervals are larger here than in previous sections as our outcomes samples are smaller.

Figure 18 presents results for low attainers, comparing those who overmatch to those who are matched to their course. The first column presents the raw difference between the two groups, akin to the difference presented in Figure 17, before conditioning on a range of demographic characteristics in the second column. The third column compares individuals who went to the same school, to remove any differences in outcomes driven by the school that the students' attended (e.g. certain schools may have better careers information which all pupils would be exposed to), while the final column also compares students within the same university type (to take account of the fact that students within Russell Group institutions are likely to have better employment outcomes, on average). This shows that:

- Low-attaining students have the same chance of dropping out whether they were matched or overmatched to their course for both our points-based and earnings-based measure of match.
- Overmatched students, for our points-based match measure, are around 6 percentage points less likely to get a 'lower class degree' than similarly low-attaining matched students, although the gap is reduced to 3 percentage points when we take into account demographic differences and differences in the schools and types of universities attended between those who overmatch relative to those who match.
- For our earnings-based measure of match, when we compare overmatched students to those who match, with similar demographics, school and university characteristics, there is no difference in the likelihood of receiving a 'lower class degree' across the two groups.

Figure 18: University outcomes by match status for low attainers (overmatched relative to baseline matched)

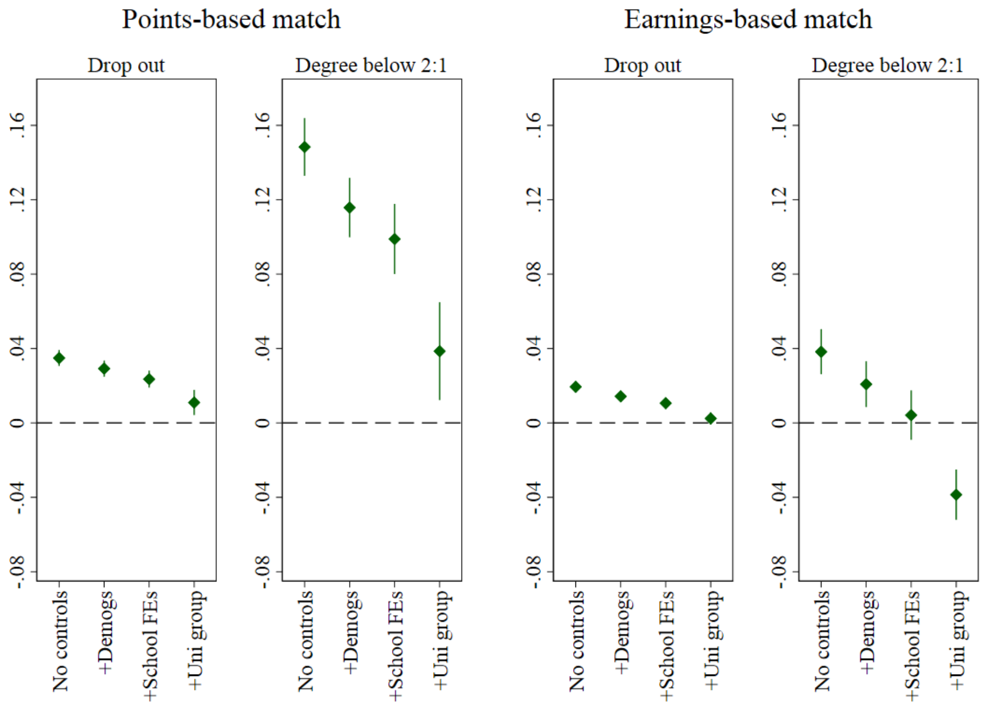


Notes on how to read these charts: the points on the chart show the difference in outcomes (drop out or degree below 2:1) of overmatched relative to matched students, with the inclusion of the controls listed on the x-axis. A coefficient of zero would mean there is no significant difference in outcomes between matched and overmatched students. A positive coefficient would mean that overmatched students are more likely to experience these outcomes than matched students, and a negative coefficient would mean that overmatched students are less likely to experience these outcomes than matched students. For example, a coefficient of -.06 on degree below 2:1 means that overmatched students are 6 percentage points less likely to get a degree below a 2:1 than matched students.

Figure 19 repeats this analysis for high attainers, comparing those who undermatch to those who are matched to their course. We can see from the graphs that:

- Undermatched students are more likely to drop out (3.5 percentage points) and more likely to achieve a ‘lower class degree’ (15 percentage points) relative to matched high-attaining students, for our points-based measure of match.
- While some of this raw difference in outcomes is explained by underlying differences in the characteristics of those who undermatch, relative to matched students (lower SES, female, White), and the schools and universities (post-92 institutions) they attend, there remains a significant penalty to undermatching for high-attaining students for both outcomes when comparing similar individuals. The main remaining difference here in observed characteristics that could explain this penalty to undermatching is that they are more likely to study social science (29% compared to 21% for matched students) or arts and humanities (28% vs 20% matched), and less likely to study STEM (40% vs 58% matched) subjects at university. We will return to this point below.
- As seen for low attainers, our earnings-based measure of match shows similar findings for likelihood of dropping out of university, although the effect sizes here are not significantly different from zero once we account for observable differences between students.
- For our lower class degree outcome, once we control for observable differences of students who undermatch relative to those who match, undermatched students are *less* likely to get a lower class degree than matched students when we measure course quality using earnings rather than achievement-based measures.
- This could be driven by the different subject composition of those who are undermatched on the points-based compared to earnings-based measures of match, with points-based undermatched students more likely to study social science and earnings-based undermatched students more likely to study arts and humanities subjects (see Table 10).

Figure 19: University outcomes by match status for high attainers (undermatched relative to baseline matched)



Notes on how to read these charts: the points on the chart show the difference in outcomes (drop out or degree below 2:1) of undermatched relative to matched students, with the inclusion of the controls listed on the x-axis. A coefficient of zero would mean there is no significant difference in outcomes between matched and undermatched students. A positive coefficient would mean that undermatched students are more likely to experience these outcomes than matched students, and a negative coefficient would mean that undermatched students are less likely to experience these outcomes than matched students. For example, a coefficient of 0.15 on degree below 2:1 means that undermatched students are 15 percentage points more likely to get a degree below a 2:1 than matched students.

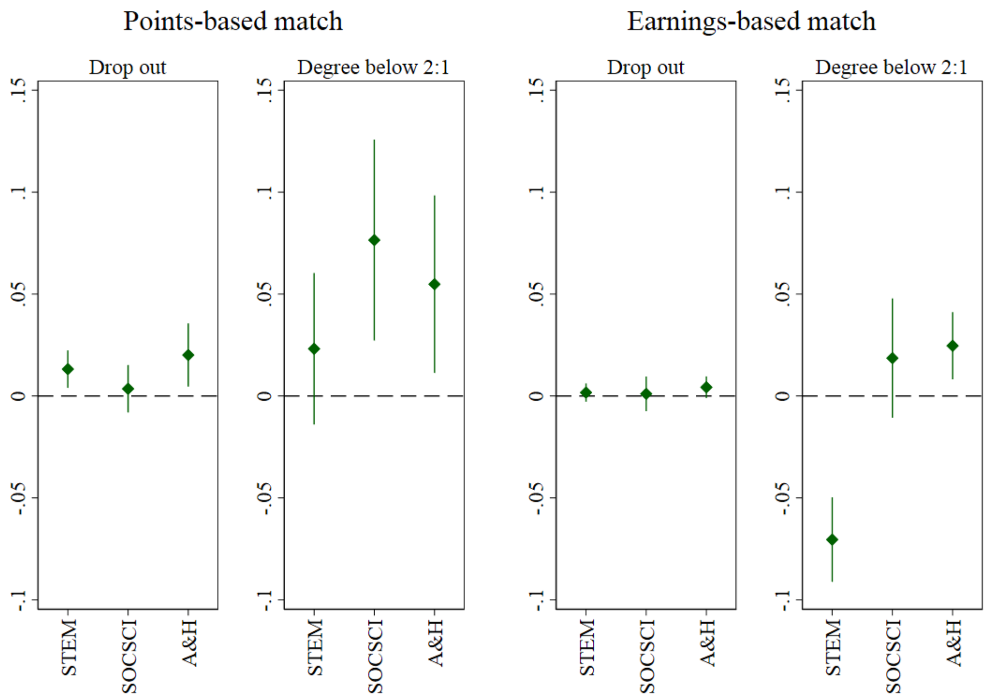
In Figure 20 we explicitly consider the role of subject studied for undermatched students, by splitting our sample by the three main subject groups. Here our results are from the final column from Figure 19, comparing individuals with similar demographic backgrounds, including gender, SES and ethnicity, from the same schools, and same university groups.

Figures 20 suggests that:

- Undermatched students, for our points-based measure of match, who are studying social science (29%) or arts and humanities (28%), rather than STEM (40%), are more likely to achieve a 'lower class degree'.
- Our finding that undermatched students are *less likely* to get a lower class degree than matched students is driven by those studying STEM, rather than social science or arts and humanities. Those studying social science or arts and humanities are actually more

likely to get a lower class degree than their matched counterparts, consistent with our findings for points-based match.

Figure 20: University outcomes by match status for high attainers (undermatched relative to baseline matched) by subject studied



Notes on how to read these charts: the points on the chart show the difference in outcomes (drop out or degree below 2:1) of undermatched relative to matched students, within the degree subjects listed on the x-axis. A coefficient of zero would mean there is no significant difference in outcomes between matched and undermatched students. A positive coefficient would mean that undermatched students are more likely to experience these outcomes than matched students, and a negative coefficient would mean that undermatched students are less likely to experience these outcomes than matched students. For example, a coefficient of 0.15 on degree below 2:1 means that undermatched students are 15 percentage points more likely to get a degree below a 2:1 than matched students, within that particular degree subject.

11 Labour market outcomes of mismatch

In our final section, we consider the labour market outcomes of those who mismatch to their university course, relative to those who are matched. As in Section 10, we focus on high and low attainers, and use a binary definition of match, comparing those who overmatch to those who match for low attainers and for those who undermatch to those who match for high attainers. The comparison group is therefore constant across our analysis but this approach allows us to separate mismatch into these two distinct types.

The differences in the characteristics of those who mismatch across our two measures of match, illustrated in Table 10, are important to remember here again: overmatched (points-based) students are typically White students studying social science at both post-92 and older non-Russell Group universities, compared to overmatched (earnings-based) students who are a slightly higher proportion ethnic minority students, studying STEM and attending mostly post-92 institutions. Undermatched (points-based) students are typically lower SES students studying social science at post-92 institutions, while undermatched (earnings-based) students are higher SES students studying STEM subjects at Russell Group institutions.

Table 11 summarises the main labour market outcomes by our alternative measures of match for low and high attainers. The table suggests that the raw differences in outcomes across match indicate that:

- Overmatched low attainers are more likely to be employed or in education, compared to matched low attainers across both measures of match. They also appear to earn more than matched low attainers.
- Undermatched high attainers are slightly less likely to be employed or in education 3.5 years after graduation for our points-based measure of match.
- For both points-based and earnings-based measures of match, undermatched students appear to earn less than high-attaining matched students.

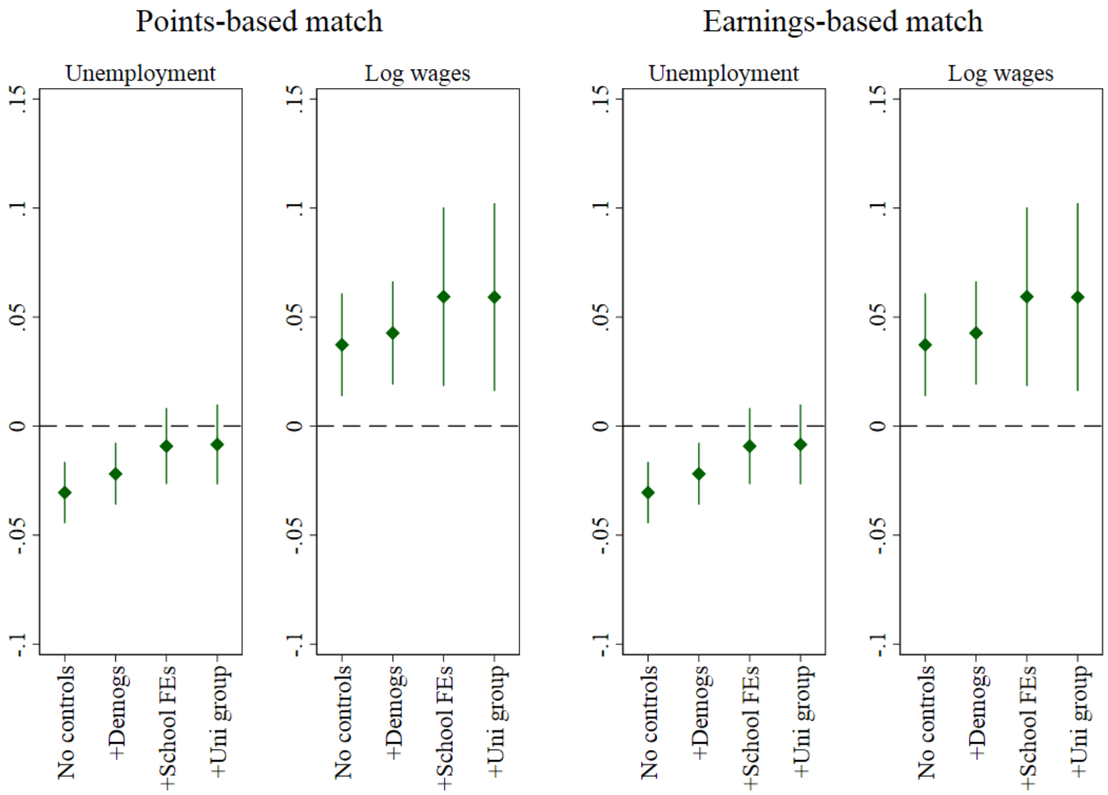
Table 11: Differences in the labour market outcomes of those who mismatch, compared to those who match, for points-based and earnings-based match

Points-based measure	*Overmatched low attainers*	*Matched low attainers*	*Undermatched high attainers*	*Matched high attainers*
Not in employment or education	10	13	9	8
Log labour market earnings	9.73	9.70	9.76	9.90
Earnings-based measure	*Overmatched low attainers*	*Matched low attainers*	*Undermatched high attainers*	*Matched high attainers*
Not in employment or education	11	13	8	8
Log labour market earnings	9.76	9.66	9.74	9.90

Given that these differences in outcomes are likely to be somewhat driven by other differences in characteristics of those who mismatch, compared to those who are well matched, Figures 21-24 present results from a regression approach, controlling for observable differences across the groups. Figure 21 presents results for low attainers, comparing those who overmatch to those who are matched to their course. The first column presents the raw difference between the two groups as in Table 11, before conditioning on a range of demographic characteristics, the school attended, and the university type. The Figure shows that:

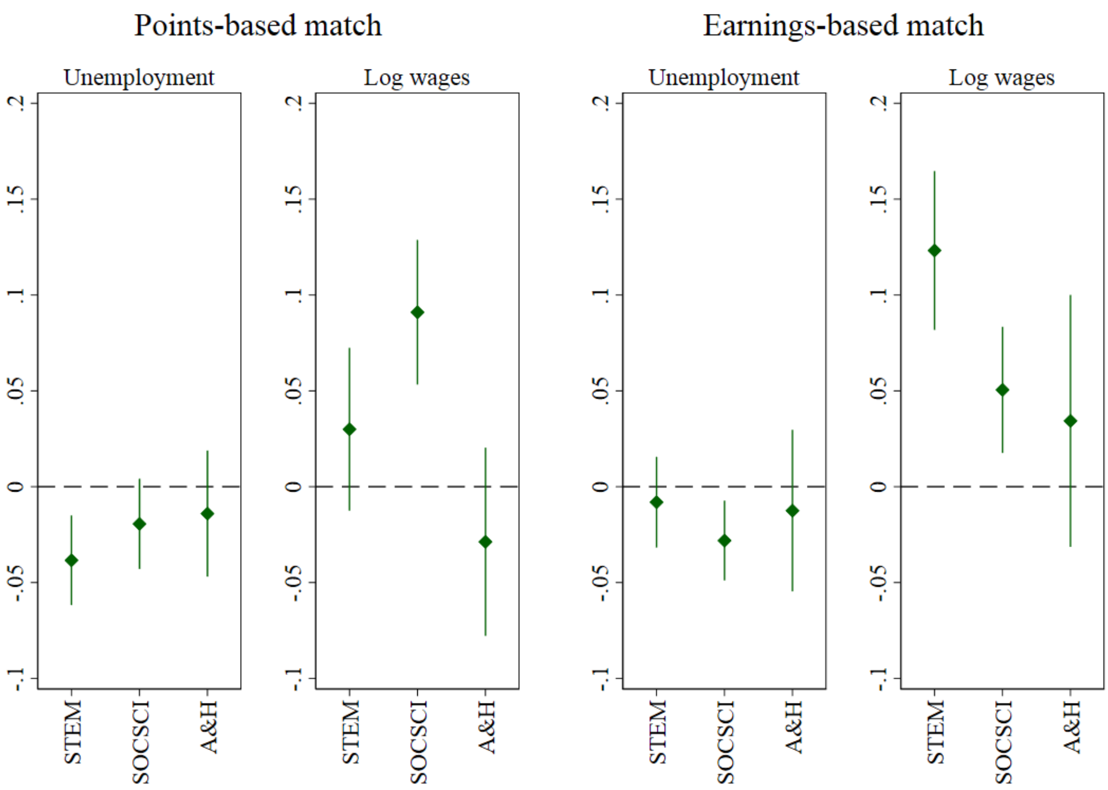
- Despite raw differences in the labour market outcomes of overmatched students, relative to matched students, for our employment outcomes, these are explained by observed differences in the school and type of university attended, rather than the fact that the student is mismatched.
- Even after controlling for a range of background characteristics, overmatched students appear to earn around 6% more than their matched counterparts across both measures of match, 3.5 years after graduation. This is primarily driven by those studying social science degrees for our points-based match and STEM degrees for our earnings-based match (Figure 22).

Figure 21: Labour market outcomes by match status for low attainers (overmatched relative to baseline matched)



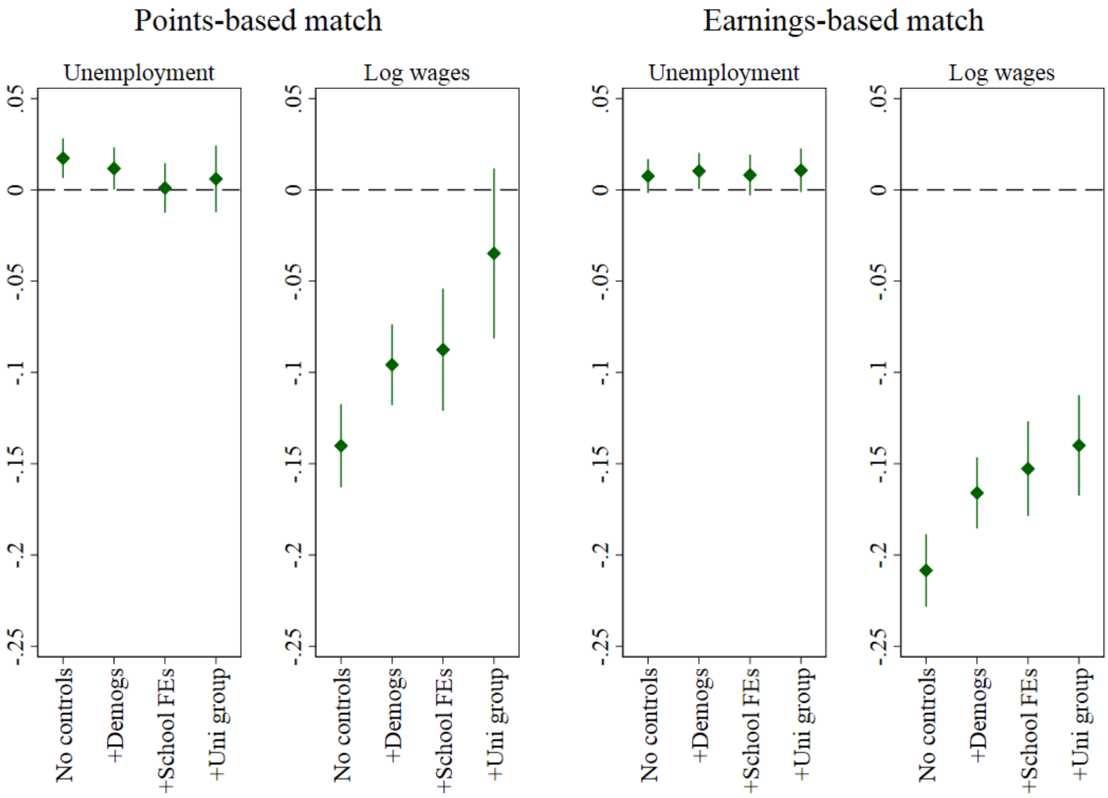
Notes on how to read these charts: the points on the chart show the difference in outcome (unemployment or log wages) of overmatched relative to matched students, with the inclusion of the controls listed on the x-axis. A coefficient of zero would mean there is no significant difference in unemployment between matched and overmatched students. A positive coefficient would mean that overmatched students are more likely to experience unemployment compared to matched students, and a negative coefficient would mean that overmatched students are less likely to experience unemployment relative to matched students. For example, a coefficient of -.03 on unemployment means that overmatched students are 3 percentage points less likely to be unemployed than matched students. For wages, a coefficient of zero would mean that overmatched and matched students on average earned the same wage 3.5 years after graduation. A positive coefficient would mean that overmatched students earn more than matched students, and a negative coefficient would mean that overmatched students earn less than matched students.

Figure 22: Labour market outcomes by subject for low attainers (overmatched relative to baseline matched)



Notes on how to read these charts: the points on the chart show the difference in outcome (unemployment or log wages) of overmatched relative to matched students, with the inclusion of the controls, by subject grouping listed on the x axix. A coefficient of zero would mean there is no significant difference in unemployment between matched and overmatched students. A positive coefficient would mean that overmatched students are more likely to experience unemployment compared to matched students, and a negative coefficient would mean that overmatched students are less likely to experience unemployment relative to matched students. For example, a coefficient of -.03 on unemployment means that overmatched students are 3 percentage points less likely to be unemployed than matched students. For wages, a coefficient of zero would mean that overmatched and matched students on average earned the same wage 3.5 years after graduation. A positive coefficient would mean that overmatched students earn more than matched students, and a negative coefficient would mean that overmatched students earn less than matched students.

Figure 23: Labour market outcomes by match status for high attainers (undermatched relative to baseline matched)



Notes on how to read these charts: the points on the chart show the difference in outcomes (unemployment or log wages) of undermatched relative to matched students, with the inclusion of the controls listed on the x-axis. A coefficient of zero would mean there is no significant difference in unemployment between matched and undermatched students. A positive coefficient would mean that undermatched students are more likely to experience unemployment relative to matched students, and a negative coefficient would mean that undermatched students are less likely to experience unemployment relative to matched students. For example, a coefficient of 0.15 on unemployment means that undermatched students are 15 percentage points more likely to be unemployed than matched students. For wages, a coefficient of zero would mean that undermatched and matched students on average earned the same wage 3.5 years after graduation. A positive coefficient would mean that undermatched students earn more than matched students, and a negative coefficient would mean that undermatched students earn less than matched students.

Figure 23 shows the same results for high attainers, comparing those who undermatch to those who are matched to their course. The results show that:

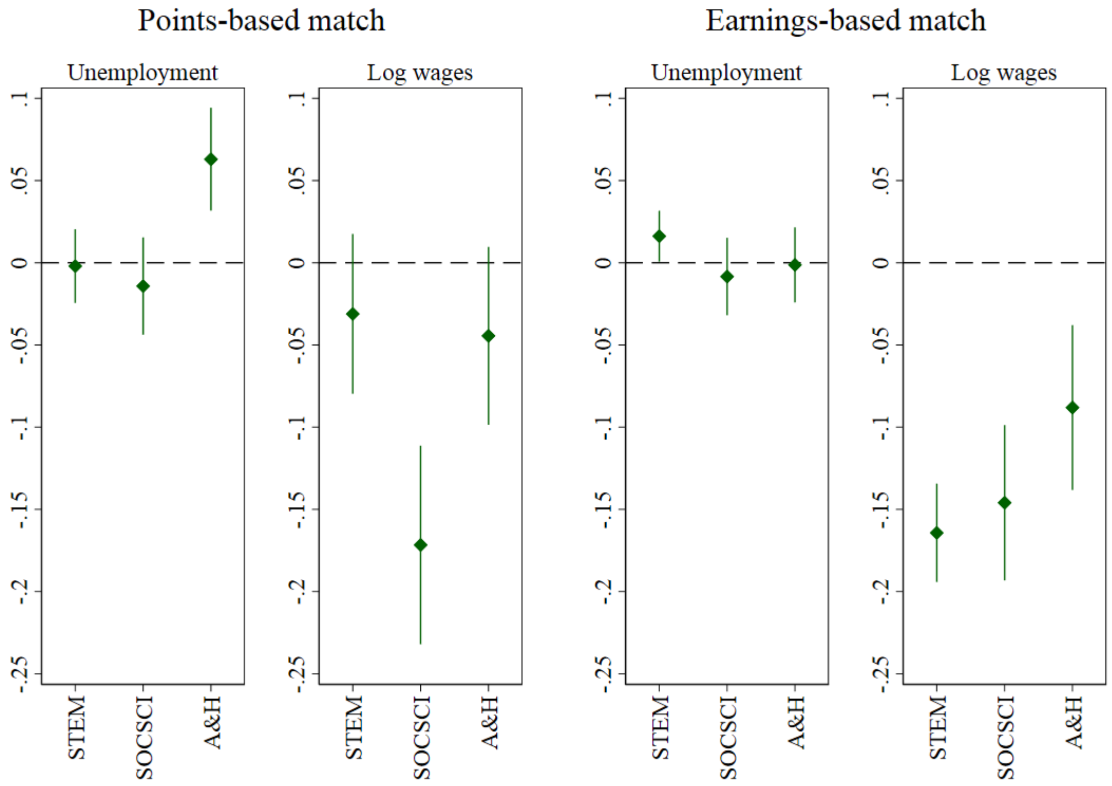
- The raw differences in employment and labour market outcomes observed in Table 11 for the points-based match are driven by differences in the schools, and particularly the type of university that undermatched students attend, compared to matched students.

- This is not surprising given the large differences in institution attended by points-based undermatched students. 74% of undermatched students attend a post-92 institution compared to 3% of matched students, whereas 70% of matched students attend a Russell Group institution compared to 7% of undermatched students.
- For our earnings-based measure of match, there are no differences in employment outcomes between undermatched and matched students.
- Undermatched students earn significantly less than matched students for our earnings-based match measure – around 15% lower earnings 3.5 years after graduation compared to matched students when comparing similar individuals who went to similar universities. Given that our measure of course quality here is labour market earnings, this is perhaps not surprising.

Figure 24 presents the undermatched results by major subject group studies at university, to explore what might be driving this wage penalty to undermatching. This shows that:

- For our earnings-based match measure, the negative wage penalty is primarily driven by those studying both STEM and social science degrees, relative to arts and humanities, although all three subject groupings have negative associations between mismatch and labour market earnings.

Figure 24: Labour market outcomes by match status for high attainers (undermatched relative to baseline matched) by subject studied



Notes on how to read these charts: the points on the chart show the difference in outcomes (unemployment or log wages) of undermatched relative to matched students, with the inclusion of the controls, by subject grouping listed on the x-axis. A coefficient of zero would mean there is no significant difference in unemployment between matched and undermatched students. A positive coefficient would mean that undermatched students are more likely to experience unemployment relative to matched students, and a negative coefficient would mean that undermatched students are less likely to experience unemployment relative to matched students. For example, a coefficient of 0.15 on unemployment means that undermatched students are 15 percentage points more likely to be unemployed than matched students. For wages, a coefficient of zero would mean that undermatched and matched students on average earned the same wage 3.5 years after graduation. A positive coefficient would mean that undermatched students earn more than matched students, and a negative coefficient would mean that undermatched students earn less than matched students.

12 Conclusions and policy implications

This project had four aims. First, to evaluate the extent of student to university mismatch in the UK, second to understand which types of students' mismatch, third, to understand the consequences of being mismatched on degree outcomes, and fourth, labour market outcomes.

We find a substantial degree of mismatch in the UK, with 15-23% of students undermatched and 15-23% overmatched, depending on the definition. While direct comparison with the US is not possible, this is suggestive that there is less mismatch in the UK than in the US, where Dillon and Smith (2017) show that around 25% of students undermatch, and 25% overmatch.

We find that low SES students are more likely to be undermatched, and less likely to be overmatched than their more advantaged counterparts; at every given attainment level, low SES students choose courses that are less academically prestigious, and that command lower earnings, than high SES students. We find a key role for secondary school attended in accounting for our SES disparities in match, with the inclusion of school effects eliminating much of the gap. This means that factors associated with secondary school such as peers, sorting, role models, and information provided by the school are the likely key drivers for improving student-to-course match.

As well as these socio-economic disparities in the quality of universities that students attend, we also find significant gender gaps. Women attend courses of similar academic prestige as men, but attend courses that command lower earnings. This is driven by subject of study; women are more likely to choose non-STEM subjects which command lower earnings. This suggests that careers-related information, such as returns to degree subjects should be targeted at female students, as well as low SES students.

We also find that White students are more likely to undermatch relative to ethnic minority students, and that students who prefer to live closer to home, who are less certain about whether they will go to university or not by age 16, and those who do not get their first choice of subject are all more likely to undermatch.

Undermatch matters for both degree and employment outcomes. There are penalties to undermatching for both university performance and labour market outcomes.

These findings have strong relevance for equity and social mobility. That high-attaining, low SES students are choosing less prestigious courses, with lower returns, has implications for their future earnings, and thus for social mobility. That high-attaining women are choosing courses with lower economic returns has implications for equity and the gender paygap.

The debate about fairness and university admissions is a live one, which these findings also have strong relevance for. The Augar Review (2019) highlighted concerns about student choice, and in particular fears that many courses are ‘low value’ and the need for increased provision of STEM courses, which are seen as more prestigious and more lucrative. Our results suggest that encouraging low SES students and women into such courses could alleviate undermatch. The review also called for a more flexible system, in which students can build up credits at lower levels which could be used to build up to a degree. While we did not look at the group of undermatched students who were qualified for a degree but instead chose a lower level qualification, such flexibility is also potentially important for this type of undermatch.

The Office for Students (who plan to announce a review of admissions imminently) have stepped up their role as a university regulator. Their role encompasses university participation, progress and outcomes, including value for money; their aim is to ensure that students participate in HE, but also that they choose the correct course. Our results suggest that there are socio-economic and gender gaps in the choices that students make, and that disadvantage should not be the only focus when considering student choice.

We offer a number of potential policy implications based on these findings.

12.1 Information, Advice and Guidance

The most obvious policy solution would be to tackle information failures affecting undermatched students. For example, students could be provided with information on the entry requirements and labour market returns to different courses at key decision-making ages (Belfield et al, 2018).

However, simply offering information (e.g. on the different returns associated with different institutions) may not be enough to resolve these issues. Studies have shown that those who gain the most from this type of information may be the least likely to consume it, and to be effective information has to be carefully targeted (McNally, 2016).

A number of trials have recently shown the positive impact of targeted interventions. A study in Michigan has highlighted the value of offering targeted information on college access and costs to high-achieving low income students who would otherwise attend a less selective college (Dynarski et al., 2018).

Two UK-based randomised control trials have similarly found value in offering targeted information to students (at school age, though a range of ages from GCSE to sixth form), providing clear guidance on the true costs and benefits of going to university.

Sanders et al. (2018) targeted schools and colleges in disadvantaged areas which sent a low proportion of students to selective universities. Students received face-to-face, inspirational talks from current students at a selective university. These talks increased the application and acceptance rates to these universities. Sanders et al. (2017), focused on high achieving young people who went to schools which typically sent more than 20% of their high achieving students to their nearest higher education institution. The pupils received a letter written by a former student of their school, with a similar background, encouraging them to aim for a selective institution. Pupils receiving the letter were significantly more likely to apply to and accept an offer from a selective (Russell Group) university.

These trials highlight the benefits of targeting specific groups: those from low income backgrounds, but also attending schools with a history of low participation in selective institutions. Thus, interventions which specifically target the key risk factors associated with mismatch are most likely to be successful – for example targeting low income students with information about local universities that are a better match for them.

12.2 The UK Applications System

Given the characteristics of those who undermatch, it could be that our current applications system is creating some of the mismatch. As discussed in section 2, students apply to universities based on their predicted rather than actual grades. Wyness (2016) showed that high achieving low SES students were more likely to have predicted grades that *understated* their actual results.

In addition, low SES students are also typically more risk averse (Schurer, 2015), meaning that they are more likely to apply to courses that are easier to access, rather than taking a risk on courses that may be harder to get in to. Both of these information failures are potential routes for mismatch to occur.

A policy solution to minimise these issues would be to introduce post qualification admissions (PQA). Creating an admissions system based on observed rather than predicted grades at A level would eliminate the issue of under-predicting for low SES students and reduce risk aversion issues, as the decision would be based on real information. This would enable all students to match more effectively to courses.

12.3 Suggested intervention

Building on these policy solutions, we propose an intervention providing targeted information within the UCAS admissions system.

Based on the idea of targeted advertising, as used by many popular websites such as Amazon and Facebook, the intervention would offer students course suggestions based on their (predicted, or preferably, actual A level (or equivalent)) subjects and grades. This system could offer a range of filters such as degree subject preference (where students would pick their preferred subject, and would be offered suggested related matched courses), and location preference (where students would be offered suggested matched courses in the area of their choice).

While this is a relatively late intervention, we are focused here on the intensive margin, improving the match between students and courses rather than encouraging participation (the extensive margin), and therefore the timing is appropriate for the desired outcome. This intervention would provide targeted information, advice and guidance, and, if coupled with PQA, could improve the quality of student to course match for those most at risk of mismatch.

Important gains in higher education participation among disadvantaged students have been made in the past few decades, with record numbers of low SES students now attending university. However, as our work highlights, there are still substantial SES gaps in the quality of the courses that students attend, with implications for equity and social mobility. Policymakers should not lose sight of the fact that it isn't just whether students go to university that matters, but where they go, and what they study.

13 References

Arcidiacono, P., Esteban A., Coate,P., and V. Joseph Hotz. "Affirmative action and university fit: Evidence from Proposition 209." IZA Journal of Labor Economics 3, no. 1 (2014): 7.

Arcidiacono, P., & Lovenheim, M. (2016). Affirmative action and the quality-fit trade-off. Journal of Economic Literature, 54(1), 3-51.

Augar, P. (2019). Post-18 review of education and funding: independent panel report. London: Department for Education.

Belfield, C., Britton, J., Buscha, F., Dearden, L., Dickson, M., van der Erve, L., Sibieta, L., Vignoles, A., Walker, I., & Zhu, Y. (2018). The impact of undergraduate degrees on early-career earnings: Research report: November 2018.

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. *The American Economic Review 105*(5), 604–610.

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

Boliver, V., Crawford, C., Powell, M., & Craige, W. (2017). Admissions in context: the use of contextual information by leading universities.

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

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 176*(2), 431–457.

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

Department for Business, Innovation & Skills. (2016). Higher education: Success as a knowledge economy–White paper. Available at: https://www.gov.uk/government/publications/higher-education-success-as-a-knowledge-economy-white-paper

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.

Dillon, E. W., & Smith, J. A. (2019). The consequences of academic match between students and colleges. Journal of Human Resources, Published online before print November 13, 2019

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.

Dynarski, S., Libassi, C. J., Michelmore, K., & Owen, S. (2018). Closing the gap: The effect of a targeted, tuition-free promise on college choices of high-achieving, low-income students (No. w25349). National Bureau of Economic Research.

Guardian, The (2019). University League Tables, 2020. Available online at https://www.theguardian.com/education/ng-interactive/2019/jun/07/university-league-tables-2020. Last accessed 17/11/2019

HESA (2019). What are HE students' progression rates and qualifications? Higher Education Statistics Agency online.

Hoxby, C. M. and C. Avery (2012). The missing "one-offs": The hidden supply of high-achieving, low income students. *NBER Working Paper 18586*.

Hussain, I., S. McNally, and S. Telhaj (2009). University quality and graduate wages in the UK. *IZA Discussion Papers 4043*.

Labour Party, (2019), Labour announces radical shake up of higher education admissions system. Online, available at: *https://labour.org.uk/press/labour-announces-radical-shake-higher-education-admissions-system/*

Mcnally, S. (2016). How important is career information and advice?. IZA World of Labor.

McIntosh, S. (2006). Further analysis of the returns to academic and vocational qualifications. *Oxford bulletin of economics and statistics*, 68(2), 225-251.

Murphy, R., J. Scott-Clayton, and G. Wyness (2019). The end of free college in England: Implications for quality, enrolments, and equity. *Economics of Education Review*
Volume 71, August 2019, Pages 7-22

Murphy, R., & Weinhardt, F. (2018). Top of the class: The importance of ordinal rank (No. w24958). National Bureau of Economic Research.

O'Kelly (2019), More students are using Clearing than ever before, wonkhe online, last accessed 13th November 2019

Sanders, M., Chande, R., Selley, E (2017) . Encouraging people into university. Department for Education, Behavioural Insights Team 2017.

Sanders, M., Burgess, S., Chande, R., Dilnot, C., Kozman, E., & Macmillan, L. (2018). Role models, mentoring and university applications-evidence from a crossover randomised controlled trial in the United Kingdom. Widening Participation and Lifelong Learning, 20(4), 57-80.

Schurer, S. (2015). Lifecycle patterns in the socioeconomic gradient of risk preferences, *Journal of Economic Behavior and Organisation 119*, 482-495.

Shapiro, D., Dundar, A., Yuan, X., Harrell, A., Wild, J., Ziskin, M. (2014, July). Some College, No Degree: A National View of Students with Some College Enrollment, but No Completion (Signature Report No. 7). Herndon, VA: National Student Clearinghouse Research Center.Smith, J., M. Pender, and J. Howell (2013). The full extent of student-college academic undermatch. *Economics of Education Review 32*, 247–261.

Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, 32, 247-261.

Sutton Trust (2004). *The missing 3000: State school students under-represented at leading universities*. Sutton Trust.

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. DOI: 10.5255/UKDA-SN-7104-4

Walker, I. and Y. Zhu (2013). The impact of university degrees on the lifecycle of earnings: Some further analysis. *BIS Research Paper No. 112*.

Winston, G. and D. Zimmerman (2004). Peer effects in higher education. In C. M. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it*, pp. 395–424. University of Chicago Press.

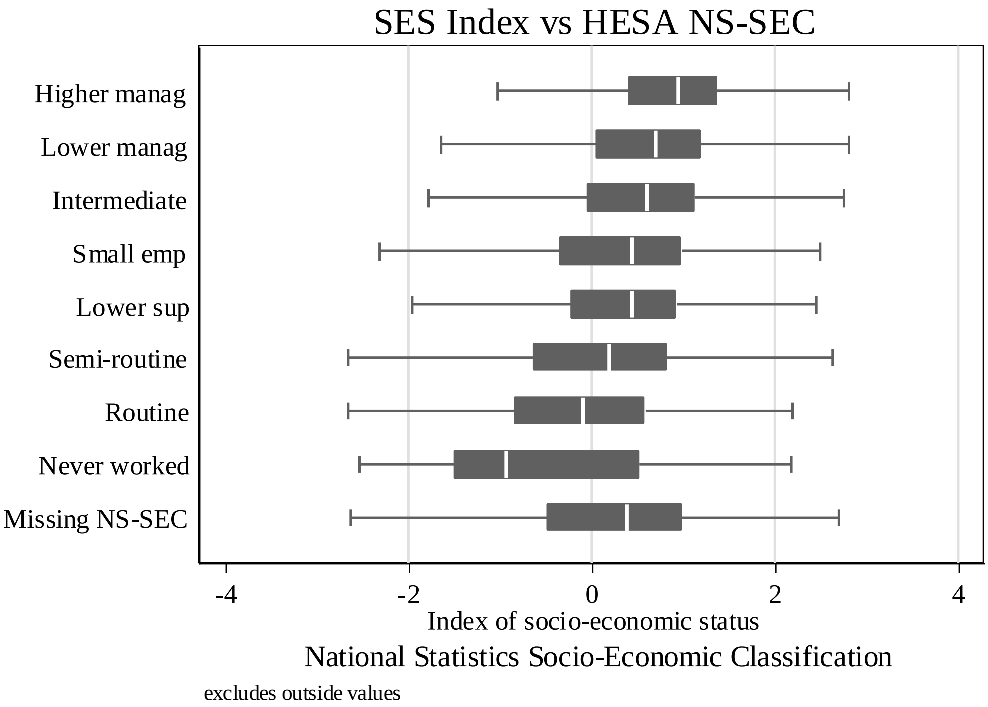
Wyness, G. (2016). Predicted grades: accuracy and impact. London: UCU.

14 Appendix

Our measures of individual attainment and course quality are based on the best three exam results. A levels are graded on a scale of E/D/C/B/A which are worth 150/180/210/240/270 QCA points respectively. Students typically study three A levels in different subjects, and the majority of universities set their entry requirements according to this measure. However a further complication is that some subjects are considered to be more rigorous than others. This is sometimes explicit, for example by naming ‘facilitating’ or ‘preferred’ subjects, and other times implicit in the offers that universities make to potential students (Dilnot, 2018). To account for these differences in universities’ subject preferences, we follow Coe et al. (2008) in calculating a subject difficulty adjustment, using an iterative approach based on our samples’ performance in different combinations of age 18 exams. For example, for students who took the same set of subjects and consistently scored higher in one of them, that subject would be deemed easier and would be awarded less points. This is iterated over all students and subject groupings until the difficultly adjusted scores are equalised.

Examples of subjects with high difficulty ratings are biology, physics and chemistry, while subjects with low difficulty ratings include film, communication and photography. We use these difficulty adjusted points when ranking students and courses in the construction of our measures of match.

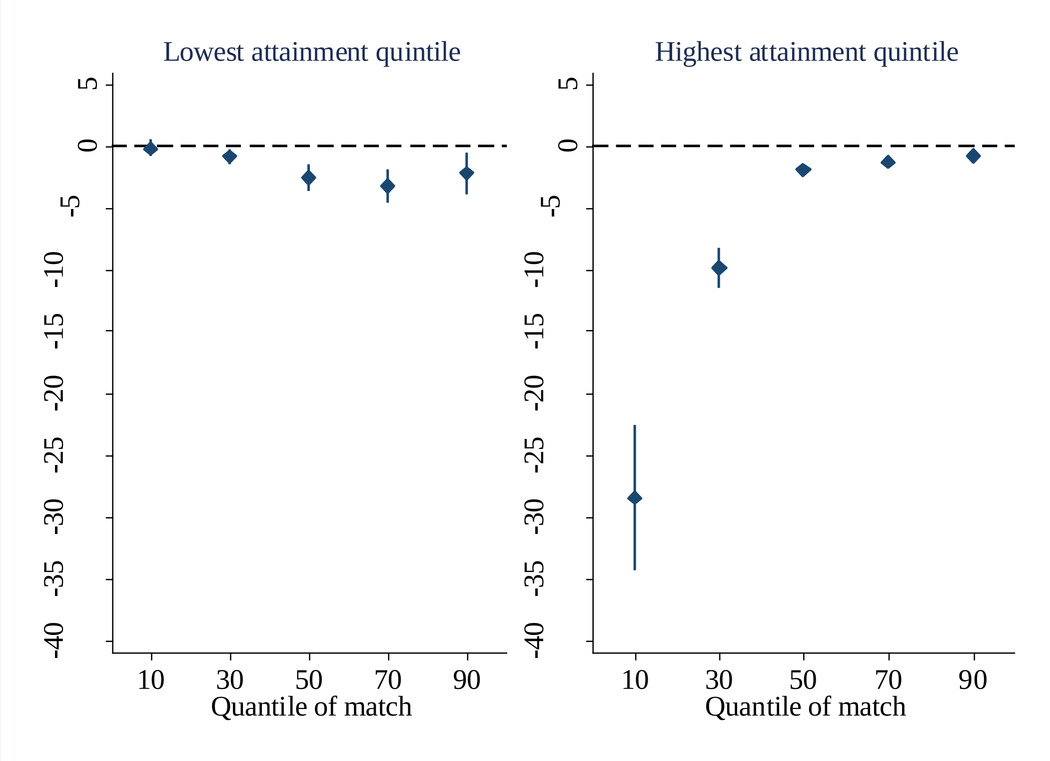
Appendix 1:Comparison of SES index with HESA NS-SEC



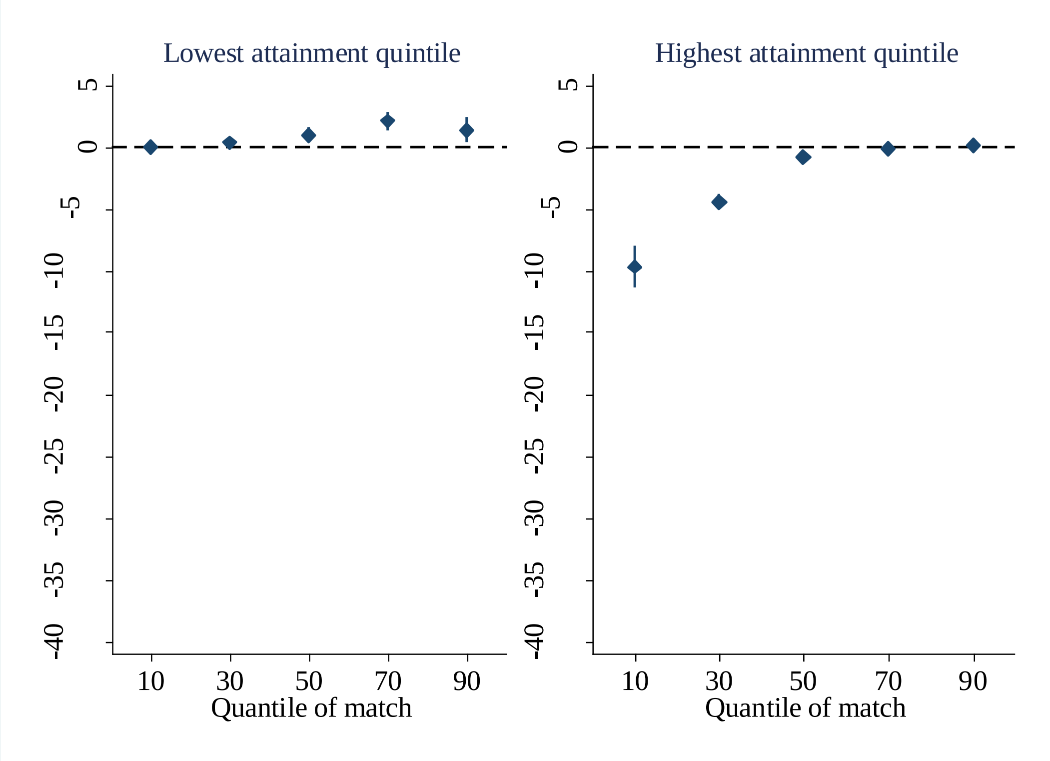
Source: NPD-HESA. n=138,969.

Figure 1 compares our SES index with the 'National Statistics-Socio-Economic Classification' (NS-SEC) which is available for around 80% of university attendees in our sample. The NS-SEC measure available in the HESA data is a fairly noisy categorical indicator of SES, since it relies on a mapping from the parental occupation which each student enters into their university application form, and has a relatively high level of non-response. Still, it is reassuring that our continuous measure of SES places the categories of the NS-SEC in a plausible ranking.

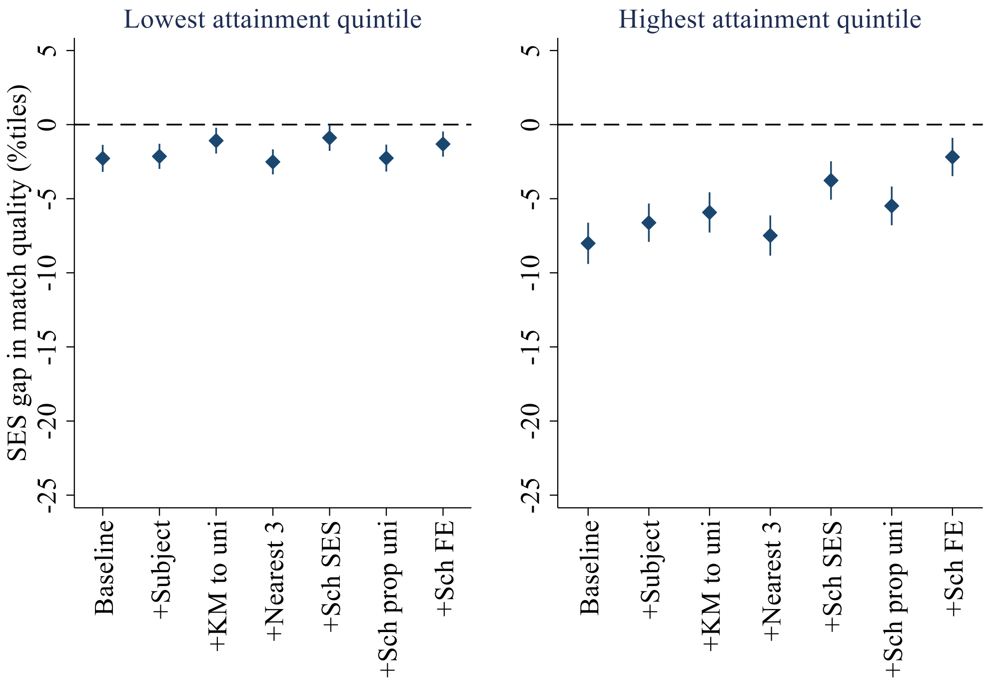
Appendix 2: Severity of match, by attainment and SES (points-based match)



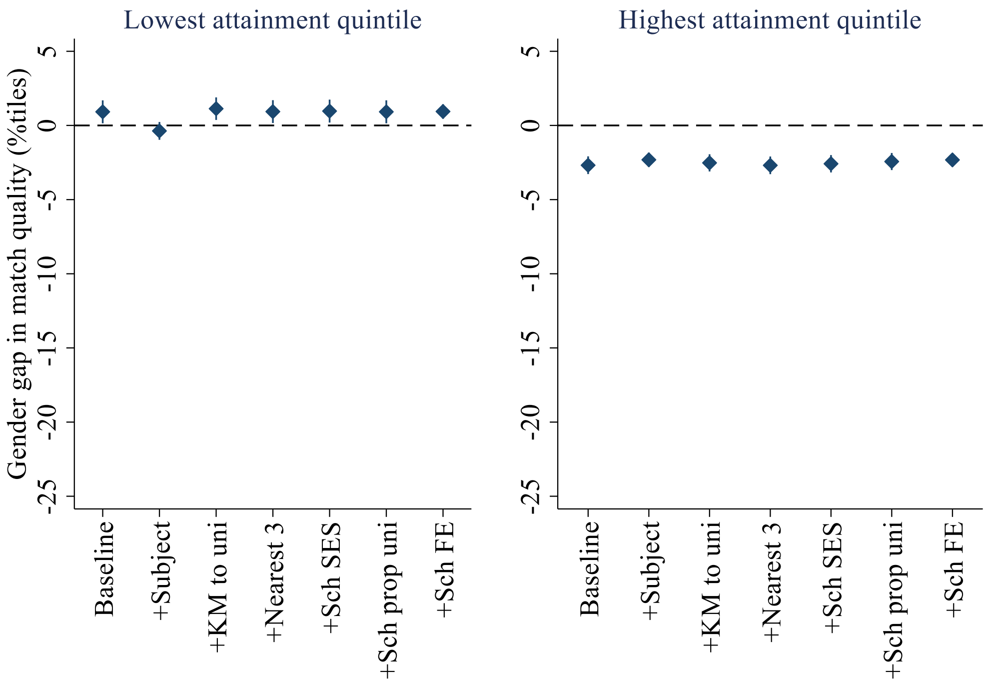
Appendix 3: Severity of match, by attainment and gender (points-based match)



Appendix 4: SES gaps in match: drivers (points-based match)



Appendix 5: Gender gaps in earnings match: drivers (points-based)



Appendix 6: Conditional SES gap in points-based match by distance to university attended (high achieving students only)