CVER Discussion Paper Series - ISSN 2398-7553

CENTRE FOR VOCATIONAL EDUCATION RESEARCH

The mismatch earnings penalty

Chiara Cavaglia, Lindsey Macmillan
Konstantina Maragkou, Richard Murphy
and Gill Wyness

Discussion Paper 041

November 2024

Centre for Economic Performance

National Institute of Economic and Social Research

LE London Economics

The University Of Sheffield.

The Centre for Vocational Education Research (CVER) is an independent research centre based at LSE, with partners in the University of Sheffield, National Institute of Economic and Social Research and London Economics. It was funded by the UK Department for Education (DfE) between 2015 and 2020.

Any views expressed are those of the authors and do not represent the views of DfE. For more details on the Centre go to cver.lse.ac.uk

Published by:
Centre for Vocational Educational Research
London School of Economics & Political Science
Houghton Street
London WC2A 2AE

The mismatch earnings penalty*

Chiara Cavaglia[1] Lindsey Macmillan[2]
Konstantina Maragkou[3] Richard Murphy[4] Gill Wyness[1,2]

October 2024

Abstract

We disentangle the relationship between student ability, course quality and the match between the two on earnings, estimating the first mismatch parameter in the literature. Using administrative data on all state-educated students in England linked to tax records, we show that high ability students attending low quality courses earn significantly less than their well-matched counterparts. By contrast, we find no evidence that lower ability students that overmatch to high quality courses go on to earn any less than well-matched students. This is evidence that affirmative action does not appear to have a detrimental effect on students' future earnings.

Keywords: mismatch, higher education, further education, returns

JEL Classification: I20, I24, I28

*We thank Susan Dynarski and Sandra McNally, the Scottish Economic Society, the Institute for Employment Research (IAB) and the LSE Centre for Economic Performance (CEP) Education work-in-progress seminar for useful comments. We acknowledge the Department for Education for funding this work through the Centre for Vocational Education Research (CVER).
[1]Centre for Vocational Education Research and Centre for Economic Performance, LSE
[2]UCL Centre for Education Policy and Equalising Opportunities
[3]Faculty of Education, University of Cambridge
[4]University of Texas at Austin
Correspondence: g.wyness@ucl.ac.uk

I Introduction

It is well established that there are substantial earnings returns to attending further and higher education (Card, 1999; Blundell et al., 2005). There is also good evidence that attending a high quality college can be a pathway to upward income mobility (Chetty et al., 2020). What is less well understood is the heterogeneous returns to course quality by student ability, or the complementarities between student ability and course quality. The recent debate on the use of affirmative action in the US (mirrored by concerns around the use of 'contextual admissions' in the UK) has raised the question of whether it is inefficient to send lower ability students to the highest quality institutions, while squeezing out higher ability students who may benefit more - the so-called "mismatch hypothesis" (Sander, 2004). This question is not yet settled, with the literature to date on the effect of affirmative action having mixed results (Bleemer, 2022; Arcidiacono et al., 2016; Arcidiacono and Lovenheim, 2016; Backes, 2012; Hinrichs, 2012).

But the study of mismatch extends beyond lower attaining students at highly selective universities (overmatch). The phenomenon of undermatch, where high attaining students attend low quality courses, is also widespread (Campbell et al., 2022; Dillon and Smith, 2017; Hoxby and Avery, 2012; Smith et al., 2013), and there is consistent evidence that undermatched students are more likely to come from disadvantaged backgrounds (Campbell et al., 2022; Black et al., 2015; Hoxby and Avery, 2012; Smith et al., 2013), making undermatch highly relevant for social mobility. Yet no paper to date has focused on the outcomes of undermatched students.

In this paper we exploit rich administrative data, which links individuals' entire education history to their tax records, to examine the earnings of both undermatched and overmatched students. Previous work (Campbell et al., 2022) has shown a significant amount of mismatch in the UK system. This may arise due to market failures such as lack of information about the benefits of attending high quality courses, preferences, or as a result of 'contextual admissions', where UK universities permit low SES students to attend high quality courses with reduced grades (Boliver et al., 2017). Previous work has shown that there are complementarities between student ability and course quality in higher education (Durlauf, 2008), implying that there is a payoff to being well-matched to one's course (Dillon and Smith, 2020; Light and Strayer, 2000). We go beyond this by illustrating the distinct relationships between undermatch and overmatch on earnings, showing the earnings for undermatched and overmatched students across the distribution of student ability and course quality. We also model the relationship between mismatch and earnings (which we find to be negative and significant) over and above student ability and course quality, for the first time in the literature.

Aside from one paper (Maragkou, 2020), the literature on the consequences of

mismatch has been solely focused on the elite group of students who attend higher education. But the majority of students across the globe do not go to university. Our analysis breaks new ground by expanding the study of mismatch to encompass a much broader set of pathways open to young people after compulsory schooling - as they move through post-compulsory education and into higher education, vocational education or go straight into the labour market. We do so by organising students into two periods of education: i) the 'post-16 stage', which occurs immediately after compulsory schooling (typically at age 17/18), when students take the qualifications to enable them to prepare for university, college or the labor market (such as SATs/ACTs in the US and A levels in England), and ii) the 'post-18 stage' - the qualification taken subsequently (usually at age 18/19), when students go on to enrol in a higher education course (such as a degree in economics at UCL), or a vocational education course (such as a certificate in plumbing at Enfield College) or have left education altogether.

While the majority of work on mismatch concentrates on the match between students and institutions, a further advancement is that (and extending methods used by Campbell et al. (2022)) we analyse mismatch at course level - defining a course as the combination of institution, subject and level of qualification. In doing so, we also break new ground by developing a measure of course quality that can be applied to a much broader range of courses than purely undergraduate degrees - including vocational courses, and postgraduate degrees.

We conceptualise the relationship between student-to-course match and future earnings as a three horse race between course quality, student ability, and the match between the two, where we are interested in which relationship dominates. The main hurdle we face is disentangling match effects from course quality and student ability effects. For example, a student would be undermatched if they are attending a course which is lower quality than might be expected given their grades. A naive comparison of the future earnings of an undermatched student and a matched student (with identical ability) might lead us to believe that there is a negative effect of mismatch. However, by definition, the undermatched student will be attending a lower ranked course than the matched student, meaning we cannot be sure that any negative effect on earnings is coming from mismatch, and is not just a course quality effect. This is similar to the common problem of trying to separate age, cohort and year effects in an earnings function.

We use three approaches to tackle this problem.

Our first is a non-parametric approach. We define student ability and course quality by ranking students nationally on the basis of their age-16 compulsory school qualifications, and by ranking courses nationally, according to the qualifications of the median student on each course. We define overmatch or undermatch as at least +/- 5 percentiles difference between student and course percentiles (students are

otherwise considered matched).

Using these definitions, we graphically examine the earnings of students who are mismatched along two dimensions - within ability: students within the same ability percentile, who attend courses of different quality (creating variation in match while holding ability constant), and within course: students on the same course who are of different ability percentiles (creating variation in match while holding course quality constant). This approach allows us to show for the first time in the literature, the earnings of undermatched and overmatched students separately, and how this relationship changes by student ability and course quality.

Our second and third approaches are parametric. For our second approach, we estimate the association between student ability and course quality (defined as above), and the interaction of the two with labour market earnings (similar to work by Dillon and Smith (2020)), showing a positive coefficient on the interaction. While this is informative in helping us to understand whether the highest ability students benefit more from attending the highest quality courses, it does not allow us to understand whether there is actually a mismatch penalty. Rather it is purely a 'complementarities parameter'.

Thus we advance on this work considerably by estimating the first 'mismatch' parameter in the literature, with our third approach. Here, we follow Hoxby and Avery (2012) and Campbell et al. (2022) by creating a match index, defined by taking the difference between the percentile ranking of the student and the course. We circumvent the problem that this 'mismatch parameter' cannot be entered into a model with course quality and student ability percentiles due to co-linearity, by entering its absolute value instead. This allows us to understand whether there is an overall mismatch penalty, over and above course quality and student ability.

From these three approaches, we are able to paint a detailed picture of the relationship between course quality, student ability, and match on earnings. We find evidence of strong positive student ability and course quality effects. However, the relative importance of ability and course quality varies substantially depending on a student's chosen pathway. For those who decide to leave education at 18, individual ability is the most important driver of earnings. However, for students who stay on and enrol in either a university degree or a further education course, course quality becomes increasingly important, almost equalising the importance of individual ability.

Moreover, the positive effects of course quality appear to significantly outweigh any mismatch effects; among higher attainers, those who overmatch to the top courses do not appear to earn any less. On the contrary, these overmatched students earn significantly more than their matched counterparts. On the other hand, we find sizeable earnings gaps for high ability students attending low quality courses. For those in the top decile of the ability distribution, attending a low quality course is

associated with as much as £8,000 lower earnings per year, compared to someone of the same ability attending a matched course. These wage gaps are far bigger for those attending university than those in FE colleges.

But are these earnings gaps purely driven by course quality or is there mismatch at play? Using the third approach we find that the absolute match parameter is negative and significant, implying that on average students that mismatch have lower earnings.

Taken together, these results suggest that in the three horse race, student ability, course quality and mismatch are all important for earnings - though the positive effects of student ability and course quality dominate the negative effect of mismatch on earnings. This suggests that affirmative action / contextual admissions policies are unlikely to cause any major inefficiencies - students attending higher quality courses than might be expected given their grades appear to benefit financially from doing so. However, our results do paint a worrying picture of the wage gap associated with high attainers attending low quality courses, suggesting policies designed to shift students to better matched courses should be directed towards encouraging low-SES high ability students to aim higher.

Our paper makes several contributions to the literature on student-to-course mismatch. The first is to add important evidence on the the distinct differences in earnings associated with undermatch and overmatch. To date, the few papers in this space have focused on examining the consequences of affirmative action (Bleemer, 2022; Arcidiacono et al., 2014), or complementarities between student and course quality (Dillon and Smith, 2020; Light and Strayer, 2000). Our high-quality data enables us to measure mismatch in more detail, illustrating the earnings of matched, overmatched and undermatched students across the entire ability and course quality distribution. Our focus on the earnings associated with undermatch is particularly unique in the literature.

A further contribution is that ours is the first paper to examine mismatch across the full range of pathways and institution types available to students after compulsory education - including vocational education. Vocational routes are an important option for thousands of young people across the globe (Hoeckel and Schwartz, 2010), as is shown by the increasing research base in the area across several countries. These pathways are often aimed at improving opportunities for young people from lower socio-economic status (SES) backgrounds, who are less likely to achieve the academic credentials necessary to access university (Chowdry et al., 2013). Our results suggest that the wage gaps for mismatch we observe largely arise from the university sector, where returns to highly prestigious courses are significant. These findings emphasise the 'high stakes' nature of higher education; students may pay a significant price for enrolling in a less selective course than might be expected given their grades. Given that low SES students are more likely to undermatch, and less likely to

overmatch, consistent with (Campbell et al., 2022; Dillon and Smith, 2017; Hoxby and Avery, 2012), it follows that the higher education sector plays a major role in generating inequality and immobility. These findings emphasise that simply encouraging low SES students to enter higher education is not sufficient to improve intergenerational mobility, and underline the importance of interventions aimed at encouraging disadvantaged students to target more selective universities (Dynarski et al., 2021; Hoxby and Turner, 2015).

II Data

Our aim is to measure the match between student ability and course quality.

We use linked administrative data from schools (through the National Pupil Database or NPD), colleges (through the Individualised Learner Record or ILR), universities (through the Higher Education Statistics Agency or HESA), and tax authorities (through Her Majesties Revenue and Customs (HMRC) tax records) in England. This linked dataset provides us with detailed information on the whole population of state-school students as they move through compulsory schooling, into further and higher education, and into the labour market.

Institutional Setting and Pathways

We consider four cohorts of students who reached the end of compulsory education at age 16 between 2002 and 2005. At this point all students take high-stakes GCSE exams in 8-10 subjects. We use students performance in this uniform assessment metric as a measure of baseline ability.

Though GCSEs mark the end of compulsory schooling for these cohorts, the majority of students remain in some form of education or training, and obtain at least one additional qualification. Acknowledging that many students do not all follow the academic path, we consider mismatch to both vocational and academic qualifications[1], organising our data into two main 'spells': i) post-16, i.e. the qualification usually taken immediately after GCSEs, and then ii) post-18, which we define as the highest qualification up to age 25.[2]

A further complication is that qualifications come in different levels, from GCSEs (a level 2 qualification), to university degrees (a level 6 qualification), with academic and vocational qualifications found in all levels.[3] We take account of this when calculating each young persons' course, defining a course as the interaction of the

[1]indeed it is possible to enter university on the basis of vocational qualifications - or to study academic qualifications then go to a FE college.

[2]Although most students have only achieved two main qualifications by age 25, around 7.5% of them have three or more qualifications. For this subgroup we aggregate qualifications together and, all other things being equal, we prioritize the one at the higher level.

[3]For more information see: https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels

type of qualification (vocational or academic), the level, the institution attended and the subject of study.

The NPD-ILR element of our linked dataset contains full information on the subjects, qualification types and levels of both vocational and academic qualifications. The HESA element of our dataset contains information on the university course of every student who is enrolled in a higher education establishment in the UK[4]. Our earnings outcomes come from Her Majesties Revenue and Customs (HMRC) tax records, including both employees and self-employed individuals. These are then matched with education registers by officials at the Department for Education (DfE) in England. We utilise the latest accessible record from 2017 which allows us to observe earnings at age 31 for the 2002 cohort, through to age 28 for our 2005 cohort.

We begin with just over 2,500,000 students. We exclude individuals who did not achieve any additional qualification after GCSEs and those with missing information on their post 16 qualifications (558,732)[5]. This leaves us with 1,845,532 students that we consider in our analysis.

Table 1: Educational pathways by ability

Post-16	Vocational	Academic	Vocational	Academic	Vocational	Academic	
Post-18	Left Education		Vocational		Academic		Total
Low achievers	11.46	1.12	5.74	0.55	0.86	0.34	20.07
Mid-ability	16.55	9.56	9.80	3.36	6.37	14.37	60.01
High achievers	0.79	2.76	0.42	0.62	1.14	14.20	19.92
Total %	28.81	13.44	15.95	4.53	8.37	28.91	100.00
Total N	531,607	247,982	294,434	83,619	154,398	533,492	1,845,532

Table 1 shows the variety of pathways open to young people after GCSEs, both academic and non-academic, populated with data from our sample. The table groups young people according to their ability: low (high) achievers are those in the bottom (top) 20% of the GCSEs grade distribution; mid-ability students are the middle 60%, and according to the two qualifications we focus on (post-16 and highest qualification). For simplicity, we classify the highest qualification into academic (mainly consisting of higher education courses) and vocational courses.

The most 'traditional' pathway would move from GCSEs at age 15/16, to an academic qualification at the post-16 stage (usually A levels, corresponding to a high school diploma), then finishing with an academic qualification - a university degree - in the highest qualification stage. This is the Academic-Academic pathway (in the final column of Table 1). Indeed around 28% of young people take this route. However, as Table 1 shows, there are many other routes available, which are also quite popular.

[4]While our school data cover students in English high schools only, we are able to track their presence in higher education institutions throughout the entire UK.

[5]This is because we are focused on mismatch on the intensive margin, e.g among those who choose to gain a qualification.

For instance, a common trajectory for low and mid- achievers, is to do a vocational qualification at the post-16 stage, and then complete their education with a second vocational qualification (16%). A small but non-trivial number of young people choose to mix vocational and academic qualifications by taking vocational qualifications at post-16, and then going on to university. 154,398 learners (8%) follow this pathway which is evidence against the common misconception that vocational qualifications cannot lead to university.

Finally, the other option available at the end of the post 16 stage is to leave education altogether. Over 40% of young people choose this path. Most notably, 28% complete a vocational course at the post-16 stage and then leave education altogether - the same proportion of young people who follow the traditional academic pathway of A levels and a degree.

Measuring student ability, course quality and match

We calculate each students percentile in the ability distribution according achievement at age 16 (the end of compulsory schooling) separately for each cohort of students.

Calculating course percentile is more complex, since students take a huge variety of courses and qualifications after junior high school (after age 16), as discussed in Section II. We expand on the methods used by Campbell et al. (2022), who only looked at higher education courses, by defining a course as the interaction of the type of qualification (vocational or academic), the level, the institution attended and the subject[6]. For example, a course could be a degree in Economics at UCL, or a Level 2 course in Health, Public Services and Care at Enfield College[7]. Finally, we compute individual and course quality separately for each cohort. Our method provides us with a total of around 25,000 courses per cohort, (98,789 in total for our 4 cohorts). We measure course quality according to the GCSE qualifications of the median student on each of our approx 100,000 course by cohort combinations.

As illustrated, a large proportion of students leave education after their post-16 qualification. We take two approaches to dealing with this. First, we place all of these students in one single course (for each cohort) of "leavers", again measuring course quality according to the median GCSE scores of students who left education after the post-16 stage. Leavers are between the 8th and 20th percentile, the variation deriving from differences across cohorts. Our second approach is to show the analysis

[6]A-levels, an academic qualification obtained at the end of high school, typically consists of 3 or more subjects. The offer of subjects is vast and students can choose any combination. For students taking A-levels, our analysis is based on the hardest subject selected by the student. The index of difficulty is computed following (Campbell et al., 2022). However, the results are consistent if using a randomly selected subject

[7]Given the large number of institutions providing vocational courses we have aggregated the subjects to 15 categories. To give an example a Level 2 Certificate in Counselling at Enfield College would be included in the Level 2 course in Health, Public Services and Care at Enfield College mentioned in the text.

excluding leavers.

Having created measures of student ability and course quality, we next follow Campbell et al. (2022) by creating a measure of match for each student in our sample, by subtracting the student's percentile in the student achievement distribution from the percentile of their course on the quality distribution. This produces a continuous measure of match for each student, regardless of the pathway and course taken after age 16. The measure represents the distance of each student's chosen course from their position in the achievement distribution.

We define a student as matched if they are enrolled on a course within +/- 5 percentiles of their own position on the student achievement distribution. If a student attends a course that is over 5 percentiles *lower* / *higher* than their own percentile in the student quality distribution, we consider them to be undermatched / overmatched.

III Methods

Our aim is to understand the relationship between student ability, course quality and the match between the two on future earnings. We can conceptualise this relationship as a three horse race between course quality, student ability, and the combination of the two (the match) where we are interested in which effect dominates.

In the absence of a source of exogenous variation in match (for example, a ban in affirmative action policy which would lead to reductions in overmatch - for example used by Arcidiacono and Lovenheim (2016) and Bleemer (2022), the main hurdle we face is disentangling match effects from course quality and student ability effects. Were we to compare the future earnings of an undermatched student and a matched student (with identical ability), and observe that the undermatched students have lower earnings, this could be driven by mismatch (specifically) undermatch, or by the effect of the student being at a lower quality course.

Being aware of this issue, we use three different approaches to descriptively examine the relationship between student ability, course quality and match.

A. Nonparametric analysis

Our first non-parametric method allows us to graphically investigate the association between different types of mismatch and earnings. We conceptualise mismatch in two different ways: i) within ability - where students of the same ability attend different quality courses (thus allowing us to strip out ability effects and hence revealing course and match effects), and ii) within course - where students on the same course have different ability levels (allowing us to strip out the effect of course quality, hence revealing ability and match effects).

For within ability match, one can imagine three students of "average" ability -

i.e. three students who all achieved a C grade average at the end of compulsory schooling. One student attends a course populated by C grade students, and is therefore matched to their course. The second student attends a course populated by D grade students, and is therefore undermatched. The third student attends a course populated by A grade students, and so is overmatched.

For within course match, we can imagine three students that are all attending the same course, which is of "average" quality - i.e. one where the median student is a C grade. One student on this course is also a C grade student, and is therefore matched to the course. The second student is an A grade student, and is therefore undermatched. The third student is a D grade student so is overmatched.

This method allows us to compare the earnings differentials of matched versus mismatched students, holding course quality or student ability constant. We plot student achievement (for within ability match) or course quality (for within course match) percentiles against earnings for matched, undermatched, and overmatched students. Plotting the raw data in this way imposes no functional form assumptions on the data and presents the raw earnings associated with being matched, undermatched or overmatched - which will be a combination of course or ability effects and mismatch effects. This descriptive approach provides a clear picture of the earnings associated with different types of mismatch.

This approach also allows us to make some inferences about the likely size of any mismatch effect, and changes to the mismatch effect by student and course quality. Looking within lower ability students, should we observe an earnings gap for those attending high quality courses (overmatch), this would suggest any (positive) course quality effects are being outweighed by (negative) match effects, with serious implications for affirmative action policies. On the other hand, looking within lower quality courses, if we observe that high ability students (undermatched) have lower earnings than matched students, this would suggest that (positive) individual ability effects are being outweighed by (negative) match effects.

B. Parametric linear regression model - estimating 'complementarities'

In our second approach, we estimate a simple model looking at the conditional relationship between student ability, college quality and their interaction on outcomes of interest, a similar approach to that taken by Dillon and Smith (2020).

In particular, we estimate our regressions using the following specification:

$$E_i = \beta_1 A_i + \beta_2 C_i + \beta_3 A_i C_i + \alpha X_i + \gamma_s + \epsilon_i \tag{1}$$

where our main outcome variable E_i is the annual earnings of individual i at age 28-31. A_i denotes individual achievement, in deciles. C_i denotes course quality, also in deciles. $A_i C_i$ is the interaction between the two. X_i is a set of individual

characteristics and γ_s are junior high school dummies. ϵ_i is the error term that we cluster at the high school level.

Thus, β_1 and β_2 can be interpreted as the correlation between student ability / college quality and earnings, while β_3 can be interpreted as their complementarity in earnings. The inclusion of high school fixed effects in our models, along with a set of individual characteristics will similarly reduce (but not eliminate) selection bias concerns in our estimates.

The individual characteristics include gender, ethnicity, socio-economic status (in quintiles)[8] and whether English is a first language. Even with our rich set of controls, we view our estimates as primarily descriptive in nature.

We advance on Dillon and Smith (2020) in a number of ways. First, we estimate equation (1) for all students leaving compulsory education at age 16, including those who leave education, and those in all types of education, not just higher education. Second, we estimate associations with earnings for both post-16 courses and final qualification courses. A third advancement is that we are able to estimate course quality, rather than institution quality in all cases, and even for both vocational and higher education courses. This is possible because of our use of highly detailed administrative data, and also given the nature of the English education system, in which students choose their major before entering further or higher education.

C. Parametric linear regression model - estimating 'mismatch'

Our second parametric approach attempts to go beyond modelling the complementarity between student ability and course quality in earnings, to try and more directly understand the relationship between mismatch and earnings, over and above student ability and course quality. Ideally we would estimate a model comprising parameters for student ability, course quality, and our match index (defined, as above, as course quality-student quality). However, these parameters are going to be co-linear. To overcome this problem, we transform our mismatch index into its absolute value, and estimate the following:

$$E_i = \beta_1 A_i + \beta_2 C_i + \beta_3 |C_i - A_i| + \alpha X_i + \gamma_s + e_i \tag{2}$$

Which differs only from (1) above with the inclusion of the term $|C_i - A_i|$, which is the absolute value of the difference between student achievement and course quality.

[8]To construct a measure of students' socio-economic status we follow Chowdry et al. (2013) and Campbell et al. (2022). We combine individual-level information including whether a student was eligible for free school meals at age 16 (around 15 percent of students), alongside a set of neighbourhood characteristics taken from the 2001 Census. These include measures on 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 2007 Index of Multiple Deprivation. We combine these measures using principle components analysis to create a standardised index which reflects the position of each student relative to the rest of the school-cohort. We then divide this measure into socio-economic quintiles.

Here, our parameter of interest is β_3 which represents the association between the extent of mismatch and earnings outcomes. A negative value of this mismatch parameter (which is the weighted average of all over and undermatched students) would indicate that there is a cost to being mismatched.

IV Results

A. Nonparametric approach

The aim of this section is to gain a simple understanding of how earnings compare for students of the same ability on different courses (within ability match), and students on the same courses, of different ability levels (within course match).

Figure 1 presents plots of within ability and within course match for post-16 qualification and post-18 qualification. For the latter, we present our findings excluding leavers (since they cannot have a final qualification by definition).

We begin by looking at mismatch within ability (students of the same ability on different courses - left hand side panels of Figure 1). First, examining post-16 qualifications, we observe that earnings are very similar, regardless of the quality of the course attended, though throughout the distribution of ability, those on higher quality courses (overmatched) earn slightly more than those on lower quality courses (undermatched). For the highest ability post-16 individuals, a more sizeable, but still small divergence emerges between matched and undermatched students. (Note at this point there are no overmatched students because it is not possible to overmatch if of the highest ability.)

By contrast, looking within ability for post-18 qualifications, we observe strikingly different earnings profiles for matched, undermatched and overmatched students, for students in the top third of the ability distribution. High achievers who attend lower quality courses earn significantly less than their matched counterparts; at the 90th percentile of student ability, undermatched students earn as much as £8,000 less than their matched peers. This earnings gap is of greater magnitude than any of the other gaps we observe across spells and within ability, and highlights the importance of course quality for post-18 qualifications. Of course, these earnings gaps - as with those for post-16 courses - will be a combination of negative course quality effects and mismatch effects.

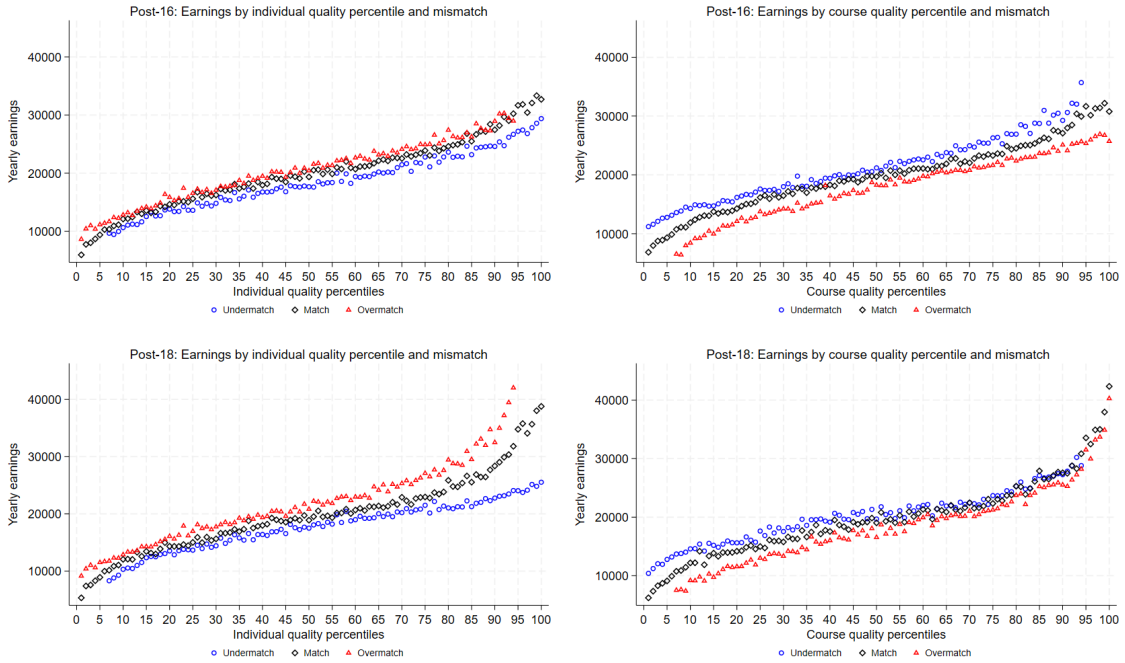
Meanwhile, we see less evidence of an overmatch earnings gap; lower ability students at high quality courses appear to earn more than their matched counterparts suggesting the positive effect of course quality outweighs any mismatch effect.

Looking within course (i.e. top right hand side panel of Figure 1), for post-16, we observe that at every point on the course quality distribution, high ability students earn more, and lower ability students earn less - particularly at the top and bottom of the course quality distribution. This suggests that student ability is the more

dominant factor for these lower level qualifications.

For post-18 qualifications, (bottom right panel of Figure 1), we see a similar pattern at the bottom of the course quality distribution. But in contrast to post-16, looking within the highest quality courses there is no evidence that overmatched students (i.e. lower ability students that may have entered such courses through affirmative action) earn any less than matched students. While there may be mismatch effects, we can infer that these are being outweighed by course quality effects. This is evidence that affirmative action (which largely takes place in the form of contextual admissions in England) does not appear to have a detrimental effect on students' future earnings.

Figure 1: Relationship between earnings, course quality and individual quality by academic match



B. Parametric linear regression model - estimating 'complementarities'

We present the results of our first estimation method (equation 1) in Table 2. Column 1 presents results for post-16 qualifications, while column 2 presents results for post-18. Column 3 reproduces results for post-18 but excludes those who leave education after their post-16 qualification.

The results confirm the presence of significant relationships between student ability / course quality and earnings, throughout the journey from compulsory schooling into further education and beyond. Interestingly, (and as was evident in the charts) the relative importance of ability and course quality varies substantially depending on a student's chosen pathway. Looking across all students in post-16 and post-18,

including those who decide to leave education altogether after post-16, individual ability is by far the most important driver of earnings. However, for students who stay on and enrol in either a university degree or a further education course (col 3), course quality becomes highly important, almost equalising the importance of individual ability.

These results suggest that individual quality is key for students with fewer qualifications, whereas course quality matters most for students with additional (and higher level) qualifications. Though not directly comparable with Dillon and Smith (2020) the results for post-18 (excluding leavers) are of a similar magnitude to their earnings estimates for 10-11 years after graduation. Our estimates are smaller, reflecting the shorter time period we have between graduation and the labour market (between 7-10 years) and the inclusion of vocational students in our models.

Our second key finding is the presence of complementarities between student ability and course quality in both post-16 (col 1) and post-18 qualifications (cols 2-3). For both spells of education, we see large and significant associations between student ability, course quality and the match between the two, on earnings. For example, for post-16 (col 1), a ten percentile increase in student ability increases their earnings by £1,068 per year, while a ten percentile increase in course quality increases student earnings by £359 per year. In line with the findings of Dillon and Smith (2020), while match matters, its importance at any stage is far outweighed by individual ability and course quality. We also observe stronger complementarities at the post-18 stage than for post-16 qualifications.

Table 2: Effect of individual quality and course quality on earnings

	Post-16 Everyone	Post-18 Everyone	Post-18 No leavers
Ability Decile	1068.002***	1045.149***	613.386***
	(21.59)	(16.11)	(23.88)
Quality Decile	359.534***	130.708***	560.787***
	(20.93)	(19.18)	(23.86)
Ability Decile × Quality Decile	34.686***	61.918***	69.276***
	(3.36)	(3.09)	(4.39)
Observations	1757608	1757608	1031720
Adjusted R^2	0.080	0.083	0.099

Notes: Clustered Standard errors in parentheses.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
The outcome variable in all models is earnings in 2017.

While this parametric approach is informative in allowing us to ascertain whether there are complementarities between student ability and course quality, it does not allow us to look at mismatch or to tell us whether there is a mismatch penalty; whether being mismatched is harmful to earnings.

C. Parametric linear regression model - estimating 'mismatch'

In this final section, we estimate a 'mismatch' parameter - β_3 from specification 2 - the effect of being mismatched to a course on earnings, accounting for student ability and course quality. The results of these estimations are presented in Table 3. The results of this exercise paint a similar picture to Figure 1: that there are sizeable student ability and course quality effects, and that the importance of course quality increases as students spend more time in education, to the point where course quality effects are as important as ability effects. For post-16 qualifications (column 1), a 1 decile increase in ability is associated with a £1268 increase in wages, while a 1 decile increase in course quality is associated with only a £543 increase. Looking at post-18 qualifications for those who stay on in education (column 3) we can see that the magnitude of the coefficient on student ability is similar to post-16 (£1009) while the coefficient on course quality has doubled to £938.

The key finding of this table, however, is the significant, negative mismatch coefficient. This indicates there is a mismatch penalty on earnings, conditioning on student ability and course quality. For students who stay on to a post-18 qualification, a one decile increase in mismatch is associated with a reduction in wages of £324. For someone earning £30,000 a year, this would be a 1% reduction in yearly wages. The mismatch effect represents around a third of the magnitude of course quality or student ability. This negative parameter is a mean of overmatched and undermatched effects. Nevertheless, the negative effect establishes that there are productivity gains from policies that attempt to match students to the appropriate courses. To our knowledge, the first time that such a parameter has been estimated in the literature.

Consistent with there being stronger complementarities in student-course match found in 2, we see that the mismatch parameter is around half the size for post-16 qualifications (col1) than for students who do post-18 qualifications (col 3). This implies our attention should be focused on mismatch at post-18.

Table 3: Student ability, course quality and absolute match. "Course rank in deciles – Individual rank in deciles"

	Post-16 Everyone	Post-18 Everyone	Post-18 No leavers
Ability Decile	1268.355***	1345.057***	1009.326***
	(13.27)	(11.40)	(11.83)
Quality Decile	543.460***	496.537***	938.712***
	(11.64)	(8.24)	(11.83)
Abs match score	-185.300***	-127.486***	-324.486***
	(15.59)	(12.83)	(16.54)
Observations	1757608	1757608	1031720
Adjusted R^2	0.080	0.082	0.099

V Discussion

In this paper we provide new evidence on the earnings outcomes of undermatched (high ability students at low quality courses), and overmatched (low ability students at high quality courses) students, compared to their matched counterparts. We observe students at two important points in their educational careers - in their first qualification after post-compulsory schooling, when they are taking the qualifications necessary to prepare them for further or higher education, and in their post-18 stage, when they are in further or higher education, or in the labor market.

Our nonparametric approach allows us to provide explorative analysis of where the earnings gaps associated with mismatch occur across the distribution of student ability and course quality. The largest gap occurs for high ability students attending low quality university courses. Such students on average earn £8,000 less per year (at aged 28-31) - or about 25% less than high ability students attending courses that are well-matched to their ability. In contrast, within the highest quality post-18 courses there is no evidence that mismatched students earn any less than matched students. This is evidence that affirmative action (which largely takes place in the form of contextual admissions in England) does not appear to have a detrimental effect on students' future earnings. Neither student ability nor mismatch appear to matter for students at the highest quality post-18 courses.

In both stages, we find evidence of complementarities between students and their courses, over and above the student's own ability and course quality. We find these complementarities to be more important for post-18 - when students make key decisions about whether to do a degree at a university, or pursue a further education qualification at college.

We go on to estimate the first mismatch parameter in the literature. We find a meaningful reduction in earnings associated with students attending courses which are not aligned with their abilities. But this negative association is far outweighed by the positive associations between course quality and student ability and earnings.

Given the consistent body of evidence suggesting that low SES students are more likely to undermatch and less likely to overmatch (Campbell et al., 2022; Hoxby and Avery, 2012; Dillon and Smith, 2017), interventions that improve the match between students and post-18 courses, such as centralised university admissions systems, or the publication of minimum entry requirements, would generate both equity and efficiency gains. The results also suggest that there would be large equity gains from affirmative action programs which traditionally move lower attaining, disadvantaged students towards higher quality courses. Moving students towards courses at which they are overmatched could nevertheless come at an efficiency loss, if there are capacity constraints which would cause the displacement of better-matched students.

Our results also emphasise individual ability is far more important than course quality for students with fewer qualifications, and those who leave education at the

end of the compulsory stage, who are likely to be the most disadvantaged students. Therefore the biggest equity and efficiency gains overall will come from improving the educational attainment of these students earlier in the pipeline.

References

Arcidiacono, P., E. Aucejo, P. Coate, and V. J. Hotz (2014). Affirmative action and university fit: Evidence from Proposition 209. *IZA Journal of Labor Economics 3*(1), 1–29. Publisher: Springer.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2016). University differences in the graduation of minorities in STEM fields: Evidence from California. *American Economic Review 106*(3), 525–62.

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

Backes, B. (2012). Do affirmative action bans lower minority college enrollment and attainment? Evidence from statewide bans. *Journal of Human Resources 47*(2), 435–455. Publisher: University of Wisconsin Press.

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

Bleemer, Z. (2022). Affirmative action, mismatch, and economic mobility after California's Proposition 209. *The Quarterly Journal of Economics 137*(1), 115–160. Publisher: Oxford University Press.

Blundell, R., L. Dearden, and B. Sianesi (2005). Measuring the returns to education.

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

Campbell, S., L. Macmillan, R. Murphy, and G. Wyness (2022). Matching in the dark? inequalities in student to degree match. *Journal of Labor Economics 40*(4), 807–850.

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

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2020). Income segregation and intergenerational mobility across colleges in the United States. *The Quarterly Journal of Economics 135*(3), 1567–1633. Publisher: Oxford University Press.

Chowdry, H., C. Crawford, L. Dearden, A. Goodman, and A. Vignoles (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 176*(2), 431–457. Publisher: Wiley Online Library.

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. Publisher: University of Chicago Press Chicago, IL.

Dillon, E. W. and J. A. Smith (2020). The consequences of academic match between students and colleges. *Journal of Human Resources 55*(3), 767–808. Publisher: University of Wisconsin Press.

Durlauf, S. N. (2008). Affirmative action, meritocracy, and efficiency. *Politics, Philosophy & Economics 7*(2), 131–158. Publisher: SAGE Publications Sage UK: London, England.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review 111*(6), 1721–56.

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics 94*(3), 712–722. Publisher: The MIT Press.

Hoeckel, K. and R. Schwartz (2010). Learning for jobs oecd reviews of vocational education and training. *Austria: Organisation for Economic Co-operation and Development (OECD)*.

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

Hoxby, C. M. and S. Turner (2015). What high-achieving low-income students know about college. *American Economic Review 105*(5), 514–17.

Light, A. and W. Strayer (2000). Determinants of college completion: School quality or student ability? *Journal of Human resources*, 299–332. Publisher: JSTOR.

Maragkou, K. (2020). Socio-economic inequality and academic match among post-compulsory education participants. *Economics of Education Review*.

Sander, R. H. (2004). A systemic analysis of affirmative action in American law schools. *Stan. L. Rev. 57*, 367. Publisher: HeinOnline.

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