

Returns to Higher Education and Dropouts: a Double Machine Learning Approach*

Sarah McNamara[†]

ZEW – Leibniz Centre for European Economic Research

April 7, 2021

Abstract

This paper provides estimates of the short-term individual returns to Higher Education (HE) in the United Kingdom, focusing on the effects of attending HE on the labour market outcomes for dropouts. Results show differential labour market outcomes for dropouts vs. individuals who have never attended HE, where outcomes are employment, wages and occupational status. I find that female dropouts, on average, have a higher occupational status than those who have never participated in HE, but do not experience a wage premium. Conversely, male dropouts experience a wage premium relative those who have never participated in HE, but the effect on occupational status is comparatively small. The evidence is mixed, however, as both male and female dropouts are more likely to be unemployed, though the effect is larger for males.

Keywords: university education, higher education, graduation, dropout, returns to education

JEL Classification: I23, I26, J31

*This research was partially conducted during my Master's thesis at the University of Mannheim in addition to ongoing work at the ZEW. I thank Friedhelm Pfeiffer for his ongoing support and expert advice, and Guido Neidhöfer, Laura Pohlen, Ralf Wilke, and Thomas Zwick for comments that greatly improved the manuscript. All remaining errors are my own.

[†]Electronic address: Sarah.McNamara@zew.de

1 Introduction

This paper estimates the short-term returns to Higher Education (HE) in the United Kingdom (UK) for a cohort of new graduates, focusing on differences in labour market outcomes between those who never attended some form of tertiary education, those who dropped out without completing a qualification, and those who completed their course of study. Thus far, the literature on labour market returns for dropouts is relatively sparse, and most studies focus on the effects of graduation. However, a partial HE-experience may be beneficial both to the individual and in terms of public investment in HE if participation is rewarded in the labour market even in the absence of a completed qualification.

For the purpose of this study I focus on individual returns, and use Double Machine Learning (DML) techniques to compare differences in the relative unemployment rates, occupational status⁷, and hourly wages between university dropouts and graduates, and those with no HE experience who possess an entry qualification. The use of DML techniques for causal inference is a relatively new area of methodological research in economics; particularly when referring to the estimation of average treatment effects (ATE) (see Athey, Imbens & Wager, 2018; Chernozhukov & Chetverikov et. al., 2018; Farrell, 2015), and even fewer studies extend these techniques for multivariate analysis in an empirical setting¹. Though there exists a body of theoretical and empirical research concerned with measuring returns to education, whether the focus be on individual, social or fiscal gains, there are complex methodological issues associated with recovering the causal effects of qualifications on later life outcomes. The use of DML techniques to provide better estimates of returns to HE is thus an interesting application, one which demonstrates how these methods can be combined with standard empirical practices for economists, and should also be of practical relevance to both policymakers and young people making the study decision and dropout decision.

I find that, on average, individuals who drop out of HE are less likely to be employed than both graduates and those with no HE experience, though the difference is more pronounced for males. In terms of wages, male dropouts who are employed command an hourly wage premium of 8%, on average, compared to their peers with no HE experience. This wage premium does not exist for female dropouts, however, whose wages are statistically indistinguishable from those with no HE experience. On the other hand, female dropouts have an occupational prestige score similar to their graduate peers, and are more likely to be employed in an occupation with a higher occupational status than high school graduates. In terms of the methodological approach, the

¹See Heigle & Pfeiffer (2019), who use DML with LASSO to estimate the effects of dropout on student outcomes in Germany, or Knaus (2018) who uses post-LASSO to estimate the returns to musicality in childhood.

results are robust to alternative specifications for tuning parameters in the DML setting.

The contributions of this paper are thus two-fold. First, it contributes to a new literature of DML methodology in economics and, as far as I am aware, it is one of few papers to use these methods in a predominantly empirical setting. Second, it not only adds to the literature on individual returns to education and the effects of dropout, but does so in a dose-response scenario where multiple “treatments” are possible; i.e. university completion status. To this end, the Next Steps linked survey (UCL, 2018) provides the data for all dimensions of the analysis and, in doing so, provides more recent estimates of the returns to HE in the UK than many other papers in the literature. Studies utilizing Next Steps as the basis of their analysis are relatively sparse and, to the best of my knowledge, no other study uses such techniques to provide a heterogeneous analysis of the wage differential between those who have attended HE, those who drop out of HE, and those who are qualified to attend but chose not to do so.

The article proceeds as follows. Section 2 summarises the literature thus far and highlights issues with existing methodological approaches. Section 3 describes the data, Section 4 introduces the empirical setting and Section 5 presents the results. Section 6 concludes.

2 Returns to Higher Education

Returns to HE in the context of this paper are those individual returns that accrue to an individual *because* of his or her study-decision, and general consensus in the literature is that there exists a strong correlation between an individual’s selection into HE and his or her later labour market successes (see Solomon (1975), or Wise (1975)). Much of the literature on drop out thus far, however, focuses on the determinants of drop out. The effects of completing *some* HE on an individual’s labour market outcomes is a relatively understudied phenomenon, and while some papers examine the consequences of drop out on labour market outcomes relative to graduates (see, for example, Gesthuizen et. al. 2005; Oreopoulos, 2007), there is a much smaller body of work that examines the effects of a partial HE experience relative to those who have never entered any form of tertiary education (see Giani et al., 2019). Consequently, dropouts are often studied from the perspective of failed graduates, rather than considering whether their (brief) HE experience may be an asset.

Yet, despite relatively sparse empirical evidence, economic theory can help explain how this partial HE experience may translate into labour market returns. First, human capital theory suggests that if an individual gained skills while (briefly) attending university, they may be more productive than individuals who only possess a HE entry qualification. Productivity differences

may therefore affect labour market returns (Becker, 1962)². Second, given the highly competitive nature of HE admissions in the UK, signalling theory suggests that having been previously enrolled in HE may convey a signal to employers about an individual's abilities and productivity (see Stiglitz, 1975), even if they did not complete their qualification. Whether or not this signal is positive or negative is more contentious. If the potential employer screens candidates based on the initial transition between secondary school and HE, an initial period of enrolment may convey a positive signal (see Arrow, 1973). We would then expect returns for dropouts to be lower than those who complete their tertiary qualification, but higher than those with no HE experience.

Conversely, drop out may convey a negative signal if it signals to employers that the individual was not academically capable, or is relatively poor at persevering when confronted with difficult tasks (see Spence, 1973). This in turn may leave dropouts relatively worse off in comparison to peers who never enrolled in HE. For example, Reisel (2013) demonstrates this to be true in the Norwegian case, and drop out is associated with lower wages than for those who complete a university entrance qualification, but do not subsequently enrol in HE. In the Swedish case, and focusing on the short-term effects of drop out, Hällsten (2017) finds that relative to those who never entered HE, dropouts spend around 3 percentage points more of their time during the first 8 years post-drop out in a state of low earnings. A third potential channel, borrowing from the sociological literature, are credentialist theories (see Collins, 1971). Credentialist theories suggest that labour market returns follow from the individual receiving some sort of certificate that is controlled by certain status groups, and that it is strictly possession of these credentials, and not increased productivity, that leads to improved outcomes for graduates.

These effects are likely to differ between countries based on underlying characteristics, such as the stock of graduates, general labour market structure, and the overall percentage of dropouts. In particular, for countries where relatively few students enroll in HE, any HE experience may convey a positive signal. Similarly, in countries where a high number of students drop out any negative signal associated with dropout may be somewhat weaker. Schnepf (2015), for example, uses PIAAC survey data to demonstrate in a cross-country comparison between several European countries that the prospects for dropouts are relatively better in countries with a low proportion of graduates, and with a higher share of the workforce participating in vocational training.

These findings are consistent with those of Matković & Kogan (2014), who focus on early career labour market outcomes in the Croatian and Serbian context and show that drop out functions as a negative signal in Croatia, but has an effect consistent with human capital theory in Serbia. As of 2017, 74.4% of Serbian secondary education students were enrolled in vocational

²For a critical overview of the literature on human capital theory, see Aina et al (2019).

schools (UIS, 2020), of which the majority (around 55%) complete a 4-year vocational programme. These effects on occupational prestige are also consistent with the findings of Heigle & Pfeiffer (2019a) for Germany, who demonstrate an increase of 5 percentage points in occupational prestige for dropouts relative to those with no HE experience, but find no effect on wages. In general, the empirical literature finds that the labour market outcomes of dropouts seem to exist somewhere in the nexus between those who complete a tertiary qualification, and those who directly enter the labour market upon completion of secondary schooling.

2.1 The UK Setting

In recent years, the UK experienced an “educational expansion”, with an increasing number of individuals choosing to enter HE upon completion of general secondary schooling. The take-up rate of university places accelerated in the 1990s (O’Leary & Sloane, 2005), with sustained increases to date. There are important policy considerations associated with this expansion in HE, however, given that it has occurred against a background of significant changes involving its financing. Since the early 1990’s, the UK has undergone a shift from the funding of student fees via general taxation to a new system of student loans and fees; ones that must be repaid in monthly instalments upon leaving HE should an individual’s income exceed a set threshold ³.

Per Blundell et al. (2000), these changes in the financing of the public university sector have seen a shift in the burden of HE-related expenses, which now largely falls on students and their parents. This is true even for students who do not complete their course of study, since repayments must be made for as long as the individual’s income exceeds the repayment threshold; irrelevant of whether the qualification was completed. This is important, as students entering HE after these changes likely faced a different incentive structure both when making the study decision, and when making the drop out decision. The results of this paper are also potentially important in this regard, as despite this cost burden a partial HE-experience may still be beneficial if participation in HE is rewarded in the labour market even in the absence of a completed qualification.

In general, the evidence on labour market outcomes for dropouts vs graduates in the UK is somewhat mixed, and suggests that drop out has a differential effect on occupation, the likelihood of employment, and wages. Davies & Elias (2003) demonstrate that dropouts are twice as likely to be unemployed than otherwise comparable graduates in the year after leaving HE, consistent with the results of Johnes & Taylor (1991) who find that dropouts are more likely to experience lengthier spells of unemployment. Yet, for those dropouts who do find employment, Davies &

³Since the academic year 2016/17, tuition fees have remained around £9,000 per annum, and students are expected to take out a full maintenance loan alongside their tuition fee loan

Elias (2003) find that they generally work in fields linked to their field of study.

Partially these results may be explained by those who make the drop out decision because they are offered a job in a relevant industry, which is more likely to be the case when a tertiary qualification is not strictly necessary for the role. Practically, however, this implies that dropouts who do subsequently find employment have earnings comparable to those of their graduate peers. Here the results of Davies & Elias (2003) differ to Johnes & Taylor (1991), as the latter find a severe wage penalty for dropouts relative to otherwise comparable graduates. In general, the literature on returns to education and dropout in the UK is somewhat sparse, however, particularly in terms of short-term outcomes. This is likely due to the fact that estimating these returns can be empirically quite challenging.

2.2 Methodological Issues and Returns to Education

The HE study-decision participants face is not a singular event, rather a process of several smaller decisions made over the course of several years. Students in the UK decide on their education pathway at 15-16, when they decide whether or not to pursue a HE entrance qualification from ages 16-18. The second stage is then whether these students obtain the necessary university entrance qualification, or do poorly on their exams and fail it. The third stage is the student's decision to attend university, or to enter the labour market (with a fourth continuous stage the decision to remain in HE until completion, or to drop out). By only considering those who possess a HE entry qualification, I am able to control for differences up until this third stage.

Distinguishing between differences arising at the fourth stage is more challenging. In the context of this paper, credentialist ideas suggest a two-fold effect when it comes to dropouts; first, given no certified qualification is bestowed, dropouts should not benefit from improved labour market outcomes relative to someone with no HE experience. Second, previous credentials are not negated upon the failure to complete a subsequent credential. Thus, there should be no cost associated with drop out, relative to those who also possess a HE entrance qualification but have never entered HE. As a baseline, credentialist ideas suggest that outcomes for dropouts and those with a HE entrance qualification only should be statistically similar. In terms of distinguishing between signalling and human capital theory, this is made difficult by the fact that these theories are not necessarily mutually exclusive. It may be the case that productivity gains do have a positive effect on the outcomes for dropouts, but that this effect is dominated by a negative signalling effect, or vice versa⁴.

⁴For a comparative analysis of human capital theory vs. screening, see Groot & Oosterbeek (1994).

Nevertheless, I hypothesize that if there is a wage premium for dropouts, this likely implies a dominating human capital effect, particularly if the effect on the other two outcomes considered here (employment and occupational status) is at least statistically similar to high school graduates. If occupational status is higher, but wages are similar to those who only possess a HE entry qualification, this likely implies a positive signalling effect; whereby companies feel these individuals are suited to a supervisory role, but do not need to pay them a higher wage given a lack of formal qualification. Finally, if dropouts are worse off on all dimensions than high school graduates, this is likely due to a negative signalling effect given human capital cannot be lost

In general, much of the literature on the returns to education relies on Mincerian-type wage regressions (see Mincer, 1974) estimated using OLS and detailed family background datasets (see, for example, Dearden, 1999; Blundell, 2000). However, OLS may perform poorly in this setting for several reasons, including linearity restrictions, dimensionality issues, potential endogeneity, and heterogeneity in treatment effects. Recent developments in the literature based on treatment effect methodology offer potential solutions to some of these issues; for example, Carneiro, Heckman & Vytlacil (2011) use an IV estimator to solve issues of endogeneity when estimating returns to college education in the US⁵, while Heckman, Humphries & Veramendi (2016) use a multi-stage sequential model to estimate treatment effects for Germany, performing a specification robustness check using several alternative estimators (including OLS, IV and nearest neighbour matching).

The latter not only accounts for heterogeneity issues, but the authors use plausible arguments for identification with no endogeneity. Similarly, Chevalier & Conlon (2003) address heterogeneity concerns by making use of propensity score matching, based on the estimator derived by Rosenbaum & Rubin (1983). What is common to all of the papers here, however, is that they are subject to two restrictions; the first is an issue of dimensionality, where the number of potential covariates that can be included without sacrificing precision of the estimates is limited by the sample size. The second is an issue of pre-specification; in all cases it was necessary to pre-specify the functional form.

There have been several more recent attempts to solve these problems via ML, for example, Heigle & Pfeiffer (2019a; 2019b) also apply the methodology described in Farrell (2015) in the German setting. By using LASSO to implement their model, issues of dimensionality and the need to pre-specify the functional form are somewhat resolved. This paper aims to contribute to a new literature in this regard.

⁵Earlier examples also include Harmon & Walker (1995); Angrist & Krueger (1991); Card (1995).

3 Data and Summary Statistics

The empirical analysis conducted here is based on the Next Steps linked survey dataset (UCL, 2018), previously known as the Longitudinal Study of Young People in England (LSYPE). Next Steps is a representative longitudinal study, which follows a cohort of young people born between 1989 and 1990. The first wave of the survey was conducted in England in 2004, when it began following around 16,000 individuals aged 13-14. Participants were surveyed annually until wave 7, conducted in 2010, at which time they were 20-21 years of age. A follow-up wave 8 was conducted in 2015/16, by which time the participants were 25-26 years of age. The survey data is linked to data from the National Pupil Database (NPD), allowing documentation of pupils in-school performance measures.

3.1 Panel Attrition and Sample Size

Of the original 16,000 individuals, the response size decreased considerably through panel attrition between waves 1-8. The overall size of the sample at wave 8, after allowing for attrition, was 7,481 observations. Due to missing data on contracted working hours, it is not possible to calculate the outcome variable of interest (in the form of hourly wages) for 65 individuals, further reducing the main sample to 7,416. Though the methodological reasoning is discussed in greater detail in Section 4, rather than comparing individuals who entered HE with all other participants, the approach here is to compare those individuals with participants who obtained a university entrance qualification (in the form of A Levels or equivalent, per ISCED classification), but chose not to attend HE.

A sample is therefore extracted containing only those individuals who possess a university entrance qualification, reducing the basic sample size from 7,416 to 4,639. Of these individuals who possess a university entrance qualification, 41.93% are male and 58.07% are female⁶. What remains is the largest possible sample that can be used to conduct the analysis. The multi-treatment analysis is then based on the identification of three treatment groups: those who gained a HE entrance qualification but chose not to attend, those who initially enrolled at a HE institution but later dropped out, and those who enrolled and subsequently completed a HE qualification. The table below demonstrates the composition of the basic sample between these groups, referred to as "entrance qualification (EQ) only", "dropouts", and "graduates". Graduates are identified based on information in the follow-up survey conducted in 2015/2016,

⁶This is consistent with more recent trends; of the 801,002 individual A Level examinations sat in 2019, 440,379, or 54.98%, were completed by females (JCQ, 2019). Typically three A Level examinations are sat per student.

Table 1: Sample Composition

Basic Sample		
	n	%
graduates	2,669	57.53%
dropouts	957	20.63%
EQ only	1,013	21.84%
total	4,639	

which asks whether participants possess a first degree or equivalent (based on ISCED values).

Dropouts are identified based on several potential channels. The first is based on education transitions made from the time participants first entered HE until Wave 7, at which time participants were 20-21 years old. If students follow the standard period of study, most enter HE at 18/19 and graduate in three years at 20/21. Entry and subsequent exit from HE in waves 5 to 7 largely identifies dropouts, as in the UK there is little flexibility in the number of credits you must take per semester. Thus, the majority of students graduate in three years (with the exception of medical sciences, and language degrees requiring an additional year abroad). To account for dropout among those entering HE education after a gap year, or at a later stage of a course with a lengthier duration, I rely on the biographical information in the follow-up wave 8. Respondents were asked if they had completed a tertiary qualification, and responses can be matched to the transitions made up until wave 7.

As Table 1 shows, 78.16% of individuals who possess a HE entrance qualification subsequently entered HE and, of these individuals, 26.39% subsequently dropped out without completing a qualification. The estimation samples used are somewhat smaller than this basic sample, given the large number of controls used in the analysis for which information is not always available for all participants. In order to exclude extreme values, the smallest and largest 1% of values for working hours and hourly wages are also trimmed.

3.2 Control Variables

The Next Steps data is particularly rich in several dimensions, which makes it useful for the analysis of dropout on labor market outcomes. First, throughout the various survey waves, information was collected on the students' academic milestones, as well as socio-economic background, employment, health status, and functional measures of non-cognitive skills. Second, in the 2015/16 follow

up wave, detailed information was collected on the students' career trajectory post-secondary education; wages, spells of unemployment etc. Thirdly, not only is a wealth of information available for the early years of the survey, when the participants were children, but participants' household characteristics are also extensively documented in the follow-up wave.

The estimation strategy based on the CIA (described in detail in Section 4) requires a large set of informative control variables. In the following I present a selection of these. For a complete list, see Table A.1 in the Appendix. From the first few waves of the survey, it is possible to identify the gender, migratory background, ethnicity, and socio-economic background of the individuals at age 13-14, to include the highest parental level of qualification, the mothers age at birth, and the occupational status of each parent present in the household (whether that be the birth parent or the other parent's partner). Occupational status is measured by the NSSEC, the official socio-economic classification system. Family background characteristics provide additional information about the number of siblings present in the household by age 17-18, and whether it is a single parent household.

Given that UK schools are administrated at local authority level, additional controls for region are included, as well as an indicator for whether the individual attended an independent school; as those fall outside the remit of the local authority. Continuity measures also identify whether the student remained at the same school for both GCSE level qualifications and the HE entrance qualification. Aside from information about the individual's university participation and wages, a number of additional controls are extracted from the follow-up wave conducted in 2015-16. These include family-related controls in adulthood, such as family composition, as well as health status and employment history. Detailed employment biographies are also available, allowing for the identification of unemployment spells. This is used in place of tenure, typical of a Mincerian-type wage regression, to account for differences in experience between graduates, dropouts, and those who enter the labour market immediately after completing their HE entry qualification.

3.3 Descriptive Statistics

Of the individuals who possess a university entrance qualification, 57.53% subsequently graduated with a tertiary qualification. In terms of sample demographics, males transitioned to HE at a similar rate to females (78.29% and 77.99%, respectively); however, males drop out at a slightly higher rate (27.79% for males vs. 24.46% for females). Based on the kernel density plots in Figures 1 and 2, there is evidence of a link between an individual's socio-economic background and their

own HE participation rate.

Figure 1 uses the highest qualification level of the participant's parent(s) as a proxy for socio-economic background, given that unemployed individuals do not have an occupational status, while Figure 2 examines the distribution of occupational status among those individuals whose father is employed. For both those with a HE qualification and dropouts, I find there is an increased

Figure 1: Kernel Density: Main Parent's Highest Level of Qualification

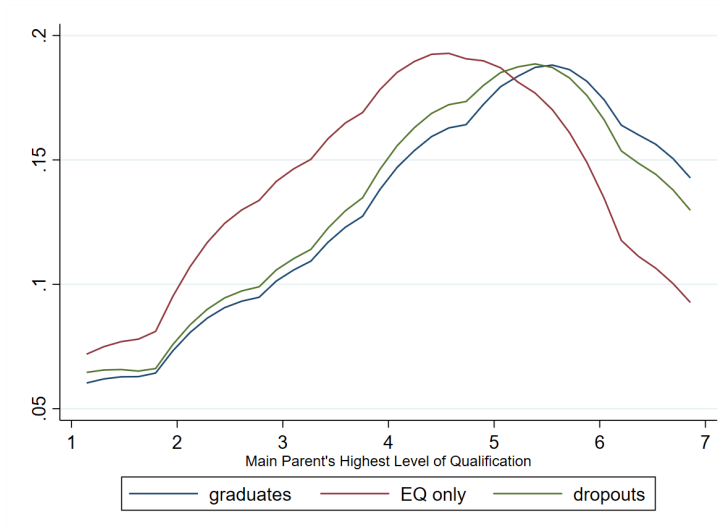


Figure 1 depicts kernel density estimates for the distribution of the main parent's education by treatment group, where 7 is "degree or higher" and 1 is "no qualifications".

Figure 2: Kernel Density: Father/Mother's Partner's NS-SEC Class

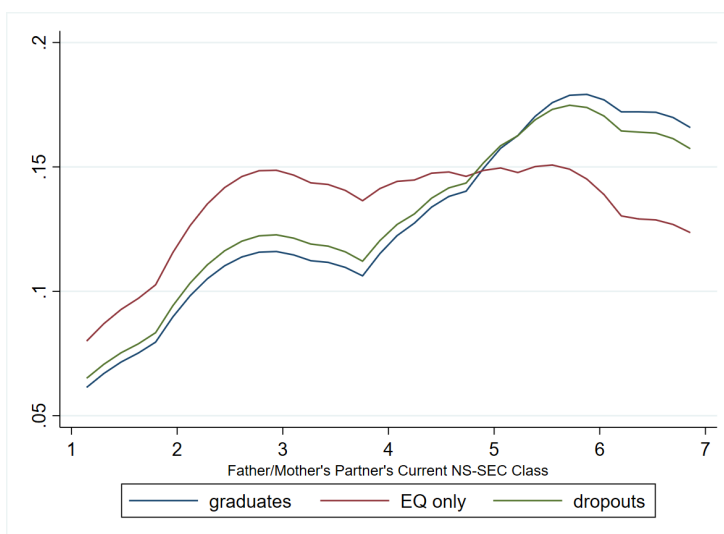


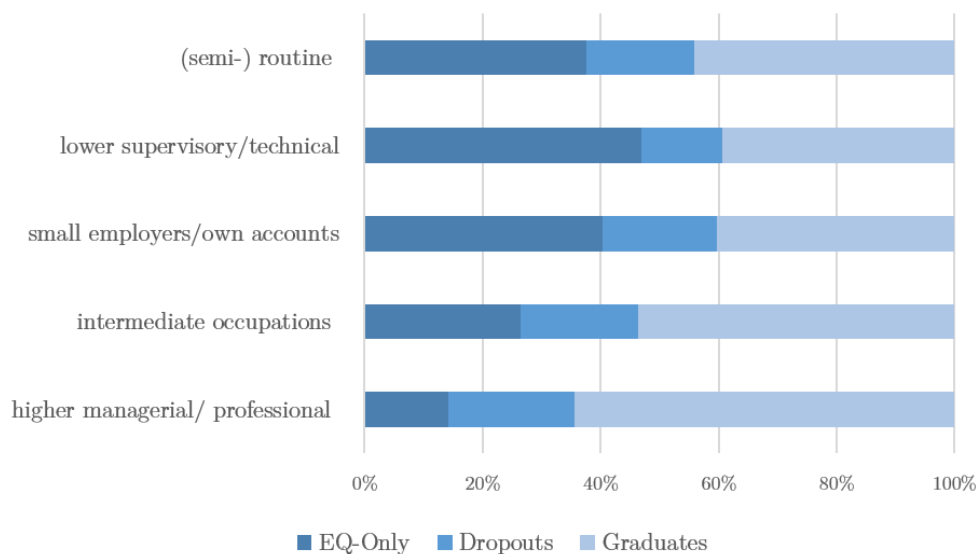
Figure 2 depicts kernel density estimates for the distribution of the father/mother's partner's NS-SEC class, where 7 is "higher managerial or professional occupations" and 1 is "routine occupations".

likelihood that their main parent has a higher level of education. The distribution is also similar for both treatment groups. For individuals who only possess a university entrance qualification, however, the main parent is likely to have a lower level of education. The relative distribution is similar when examining the occupational status of the individuals father. Empirically, graduates are more likely to have a father with a high socio-economic status, while those who have no HE experience are more likely to have a father that works in a lower supervisory or technical role (occupational score of 3).

3.4 Early Career Employment Path

Finally, in order to better understand the different destinations post-secondary education and post-HE, it is informative to examine the basic labour market outcomes between the three groups in the sample. Individuals who possess a tertiary qualification are slightly more likely than their peers to be employed; unemployment among graduates was 13.94% during the follow-up wave, rising to 15% for those with no post-secondary education, and highest among dropouts at 15.88%. This is consistent with the results of Davies & Elias (2003), who demonstrate that dropouts in the UK context are more likely to be unemployed than otherwise comparable graduates. Furthermore, per Figure 3 below⁷ a higher share of individuals who possess a tertiary qualification appear to be working in job roles with a greater degree of autonomy than their peers who possess only a university entrance qualification.

Figure 3: Distribution of Occupational Status



⁷Based on Table A2 in the Appendix.

Based on the NSSEC five analytic classes for occupational status, of the sampled individuals who are in active employment at the time of the follow-up wave, 64.49% of those working in a higher managerial or professional occupation are graduates. Conversely, 14.18% only possess a university entrance qualification. Of the 803 dropouts who were gainfully employed at the time of the follow-up wave, 519 (or 64.63%) were working in higher managerial or professional roles, with the bulk of the remainder working in intermediate roles.

The opposite is true when it comes to lower technical roles and semi-routine or routine occupations. Of the 856 individuals who only possess a university entrance qualification, 30.96% are employed in a job role with a lower occupational status, vs. 13.70% of dropouts and 12.09% of graduates. These findings are consistent with the results of Matković & Kogan (2014) for Serbia, and Heigle & Pfeiffer (2019) for Germany, who note that HE experience, even without completing a HE qualification, is associated with a higher occupational prestige score. Per Pearson’s chi-squared test, it is not possible to reject the null hypothesis of no significant statistical dependence between occupational class and HE experience. Consequently, in addition to estimating individual financial returns to HE in the form of wages, the effect of HE on occupational status is also investigated in greater detail in the following sections.

Figure 4: Average Hourly Wages by Treatment Group

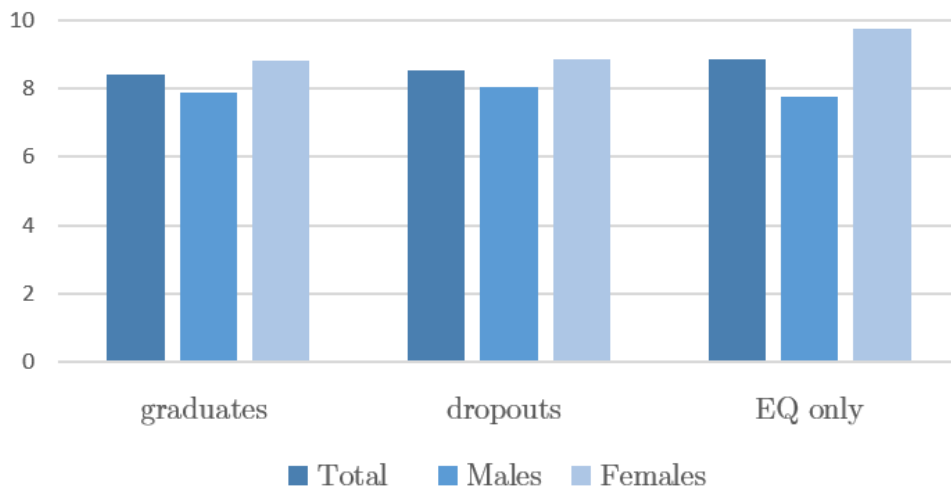


Figure 4 shows the mean hourly wages by treatment group and gender⁸. Without this further level of disaggregation graduates earn the *least*, on average, while those who only possess a HE entrance qualification tend to earn the most. Although the differential is not particularly large, at £0.45 per hour. Disaggregating by gender, however, demonstrates that male dropouts earn a similar wage to male graduates, and both earn more than their counterparts who only possess

⁸Based on Table A3 in the Appendix.

a HE entry qualification. Females have higher average hours wages initially but suffer a wage penalty with "dropout". In both cases graduates earn the least although this is likely, at least in part, due to salaried workers; if workers are salaried rather than paid an hourly rate, working more hours per week means a corresponding lower hourly wage. I return to this observed wage differential issue when interpreting the empirical results.

4 Methodology

In order to estimate the effect of the various "treatments" on the outcome variables of interest, I draw from the literature on the potential outcomes framework (see Rubin, 1974). Letting y_{ti} denote the potential outcome for individual i (i.e. log hourly wages or occupational status) had they received treatment t where y_{ti} refers to a realisation of y_t ; a random variable. Then, let y_0 denote the potential outcome of an individual who received no-treatment at all; that is, they are a non-graduate. In standard treatment analysis, y_1 would be used to denote the potential outcome under treatment (i.e. obtaining a HE qualification). Here, however, the analysis is extended to the multivariate case; thus, per Lechner (2001), I denote individual-level treatment effects as $y_t - y_0$ for a range of $T + 1$ different treatments $D_i \in \{0, 1, \dots, T\}$. Then, for each for individual i , though y_{ti} denotes all potential outcomes for each $D_i = t$, only one realisation is observed. The parameters of interest are then:

- **POM**: the potential-outcome mean for each t as an average of each potential outcome:

$$POM_t = E(y_t)$$

- **ATE**: the average effect had individual i received treatment t rather than treatment 0 :

$$ATE_t = E(y_t - y_0)$$

4.1 Double Machine Learning (DML)

In this paper, I follow the methodology proposed by Farrell (2015) whose DML estimator builds on the efficient score for semi-parametric average treatment effect (ATE) estimation proposed by Hahn (1998):

$$POM_t = \mu_t = E \left[\frac{d_i^t (Y_i - \mu_t(X_i))}{p_t(X_i)} + \mu_t(X_i) \right], \forall t \quad (1)$$

where $d_i^t = \mathbf{1}\{D_i = t\}$ denotes the treatment indicator, $\mu_t(x) = E[Y_i | D_i = t, X_i = x]$ the conditional expectation of the outcome, and $p_t(x) = P[D_i = t | X_i = x]$ the conditional probability of receiving treatment t .

Here both the outcome *and* the propensity score are treated as high-dimensional nuisance parameters; using Machine Learning (ML) techniques to jointly approximate said nuisance parameters. Using this approach, Farrell (2015) demonstrates that it is possible to obtain uniformly valid inference even when model selection is conducted post-estimation, although it was Belloni et al. (2017) who generalised this approach. They showed via their work on the moment condition identification of these parameters that it is possible to fit ML methodology into the causal inference framework, demonstrating when parameters are identified via moment conditions that in turn satisfy the orthogonality condition⁹, inference is uniformly valid. This is because the moment conditions used for identification are robust to small model mistakes when jointly approximating the nuisance parameters for propensity score and the conditional potential outcome mean (POM). This technique is thus referred to as “double” machine learning because it is two equations, jointly estimated, using a doubly robust estimator.

Per the findings of Chernozhukov & Chetverikov et. al. (2017), the above estimator fulfills the orthogonality conditions by nature of being a semi-parametrically efficient score. Taking the derivative of the moment condition with respect to $\mu_t(x)$ and $p_t(x)$, the parameters Farrell (2015) refers to as “nuisance parameters”, is 0 at μ_t ; the POM described previously. The variance is then the square of the efficient score when i.i.d is satisfied:

$$\sigma_{\mu,t}^2 = E \left[\left(\frac{d_t^t(Y_i - \mu_t(X_i))}{p_t(X_i)} + \mu_t(X_i) \right)^2 \right], \forall t \quad (2)$$

In order to obtain the POM’s based on the algorithm above, estimation must proceed as follows:

1. Predict conditional outcomes $\hat{\mu}_t(x)$.
2. Predict conditional treatment probabilities $\hat{p}_t(x)$.
3. Plug estimates for $\hat{\mu}_t(x)$ and $\hat{p}_t(x)$ into the sample analogues of the above equations.

By using the DML estimator proposed by Farrell (2015), I am able on the one hand to estimate both the POM’s and the likelihood of treatment without the need to pre-specify the functional form and, on the other hand, the results of Belloni et. al. (2017) ensure that as long as orthogonality conditions are met, results can be interpreted within the causal inference framework. Furthermore, I follow one of the two extensions to Farrell (2015) proposed by Knaus (2018). Knaus (ibid.) extends the methodology in two important ways; first, he derives a weighted representation of the estimator which can be used to assess covariate balancing and second, he proposes a data-driven assessment inspired by the One Standard Error (1SE) rule of Breiman et.

⁹For a full review of orthogonality conditions, see Neyman (1959).

al. (1984) in order to address concerns about the sensitivity of the estimators to tuning parameter choices. In this paper, I apply the second extension (see Appendix B).

4.2 Identification

Before I am able to proceed, however, it is first necessary to derive the conditions under which identification is possible. The individual’s decision to participate in HE (and to subsequently drop out), is not a random event. A number of individual characteristics related to an individual’s family and socio-economic background, personal characteristics and school-related factors may affect an individual’s decision to obtain a tertiary qualification. This makes identifying the causal effect of the ”treatment” more difficult. When selection into treatment occurs, the POM described is not equivalent to $E(y_i)$ and estimates of treatment effects will be biased. In order to get identification of this POM, a number of conditions must be met. First, there must exist a vector of observed characteristics \mathbf{X}_i that are exogenously predetermined, such that both the conditional independence assumption (CIA) and the common support assumption (CSA) are fulfilled:

- **CIA:** $y_{ti} \perp\!\!\!\perp D_i | \mathbf{X}_i = x, \forall i, t$ and $\forall x \in X$
- **CSA:** $P[D_i = t | \mathbf{X}_i = x] > 0, \forall i, t$ and $\forall x \in X$

The CIA requires that, after conditioning on the vector \mathbf{X}_i , treatment status is as good as random, as anything that may affect treatment status is accounted for. When estimating treatment effects in a potential outcomes framework, the CSA implies that non-treatable observations are not present in the data. That is, the CSA requires that for any individual i , the likelihood they receive any treatment t from the set of all possible treatments $D_i \in \{0, 1, \dots, T\}$ must be greater than zero. Thus, in order to satisfy the CIA and the CSA, only those who obtain the university entrance qualification are considered. This does have the unfortunate side effect of reducing the generalisability of the findings.

4.3 DML Estimation: Implementation of Post-LASSO

To implement this, per Knaus (2018) I use 10-fold cross-validation of the Post-LASSO to determine the penalty term minimising the out-of-sample MSE, while per Belloni & Chernozhukov (2013) separate logistic Post-LASSO regressions of each treatment indicator are used to calculate the predicted propensity score. Post-LASSO estimation is based on the LASSO, as described in Tibshirani (1996). Fundamentally, LASSO simply controls the extent to which the vector of

control variables are permitted to enter the model. That is, a penalty term λ is applied to the covariates, which either "down-weights" them, or removes them from the model entirely.

It is for this reason it is referred to as a shrinkage estimator; intuitively, if independent covariates contribute very little to the model in terms of explaining variation in the outcome parameter of interest, then they ought not to play a large role in estimating model parameters. The following is the optimisation problem that LASSO solves, where jack-knife cross-validation is used to identify the optimal λ :

$$\min \left[\sum_{i=1}^n (Y_i - X_i\beta)^2 \right] + \lambda \sum_{j=1}^p |\beta_j| \quad (3)$$

In the case that $\lambda = 0$, the LASSO collapses to the OLS estimator. The Post-LASSO implementation is then simply the un-shrunken coefficients obtained via OLS regression, with the Post-LASSO non-zero estimates (i.e. the estimates associated with coefficients that have not been dropped from the model) used for prediction. The result is an estimated POM for each "treatment", which can be used to calculate the ATE of receiving said "treatment". Treatments in this setting, of course, refer to the HE study-decision, and subsequent dropout decision.

5 Results

5.1 Employment

The sample unemployment rate for the different treatment groups indicates that individuals who possess a tertiary qualification are slightly more likely than their peers to be employed; wherein unemployment is highest among dropouts at 15.88%. As a first result, Table 2 presents estimates of the effects of the different "treatments" on the likelihood of employment, where the outcome is a binary logistic model equal 1 if the individual is employed and 0 if they are not. Table 2 is based on a low-dimensional model, consisting of the control variables described in Table A1 in the Appendix with the exclusion of job-related controls. The results were estimated both on a pooled sample and then separately for males and females, and show that, given a comprehensive set of background characteristics, graduates are less likely to be employed than those who directly enter the labour market post-secondary education. This is not unexpected, given this follow-up wave was measured only several years post-graduation, and graduates may not yet have found long-term job roles.

The results on dropouts, however, show that those who leave HE without completing a qualification are less likely to be employed than both graduates and those who only possess a HE entrance qualification. This is true for both males and females, though the differences between

Table 2: ATE’s Post-LASSO: Dropout and Employment (High Dimensional Setting)

	Pooled		Males		Females	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
ATE						
graduates	-0.0389825**	0.0135005	-0.0785582***	0.0181033	-0.023946	0.024489
dropouts	-0.0448937**	0.0168176	-0.0874307***	0.0245432	-0.034985	0.028331
POM						
graduates	0.86174	0.078	0.86134	0.0116	0.86558	0.0103
dropouts	0.85583	0.0127	0.85247	0.0202	0.85454	0.0173
EQ only	0.90073	0.0112	0.93990	0.0142	0.88952	0.0224
on/off support	3516/160		1525/31		2034/86	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

treatment groups are smaller for females. Although the POMs for male and female dropouts and graduates are similar, the POM for those with a HE entrance qualification only is much smaller for females, consistent with the idea that females withdraw from labour market for other reasons (i.e. children). That dropouts experience this additional penalty in comparison to the other two treatment groups, and the POM is similar for both males and females, offers a first hint that there is a negative signalling effect associated with dropout.

5.2 Treatment Effects on Hourly Wages

To establish a baseline, I first use a low-dimensional model consisting of the 66 variables described in Table A1 in the Appendix. The results were estimated on a pooled sample and then separately for males and females (see Table A4). In this low-dimensional setup I find a negative ATE not only for dropouts, but also on the hourly wage rate for individuals who complete a tertiary qualification. At around £0.37 per hour this graduate wage penalty, relative to those who only possess a HE qualification, is larger than the dropout wage penalty at around £0.17 per hour. Results for this low-dimensional pooled sample imply that hourly wages are highest among those who have never attended any sort of HE.

However, the above is based on a relatively low-dimensional model and only one of the coefficients is statistically significant. The benefit of the DML methodology is the ability to relax

dimensionality restrictions, and Table 3 presents results based on a high-dimensional model. In

Table 3: ATE's Post-LASSO: Dropout and Hourly Wages (High Dimensional Setting)

	Pooled		Males		Females	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
ATE						
graduates	-0.157828	0.141414	0.32568	0.17921	-0.2861990	0.2558587
dropouts	0.058639	0.163163	0.61126**	0.22048	-0.0066096	0.2767873
POM						
graduates	8.5193	0.0651	8.0359	0.0886	8.8754	0.0926
dropouts	8.7358	0.1037	8.3251	0.1566	9.1550	0.1410
EQ only	8.6771	0.1292	7.7102	0.1604	9.1616	0.2419
on/off support	2887/73		1225/44		1604/84	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

the higher dimensional setting, interactions of all of the baseline variables with ethnicity, and female (in the pooled model) are included, along with second and third degree polynomials of the continuous variables. In this case, the hourly wage rate for female dropouts is virtually indistinguishable from those who only possess a HE entrance qualification. The wage penalty persists for graduates in the high-dimensional model in the female case, although the ATE is not statistically significant. For males hourly wages are, on average, around £0.33 per hour higher for graduates, but are higher still for male dropouts, who experience a wage premium of around 8% in comparison to those with a HE entrance qualification only. The coefficient on dropout is also statistically significant at the 0.01 level.

When interpreting these results several factors should be kept in mind; first, there is some indication that dropouts, particularly in certain disciplines, appear to be working in different *types* of job role with a greater degree of autonomy. For males, if dropouts and individuals who only possess a HE entrance qualification are structurally different in terms of occupational choice, the wage differential estimated here may be indicative of a positive selection on dropouts. If these individuals have acquired functional skills relevant to their intended field during their spell in HE, human capital theory suggests that, if they subsequently enter an industry where their skills are practically relevant, they will be correspondingly rewarded in the form of higher wages.

Furthermore, if there are differences in returns to different subject areas it may be that the ATE's estimated here are indicative of a broader issue, such as the relatively poor job prospects for certain types of HE qualification and the risk of graduate underemployment. What the results here do suggest, however, is that drop out does not impose a negative penalty on income relative to those who have never participated in HE as (negative) signalling theory may otherwise suggest. This is also somewhat consistent with credentialist explanations, given I find no wage penalty; indicating that possession of the HE entrance qualification sets the lower bound for wages. Female dropouts are at least as well off as their peers with no HE experience.

5.3 Treatment Effects on Occupational Status

For occupational status, the NSSEC 5 have been coded as 1 through 5, where 1 indicates a (semi-) routine occupation, and 5 a higher level professional role. As in the previous section, I first calculated estimates using a low-dimensional model (see Table A5) consisting of the baseline variables from Table A1, with the exclusion of job-related characteristics and the addition of parental occupation. Results are estimated both on a pooled sample and separately for males and females; and show very statistically significant similarly positive effects on occupational status in the pooled sample for both dropouts and graduates, though the disaggregated models show this is largely driven by females. For females, any HE experience has a large positive effect on occupational status independent of whether the qualification was completed. Yet, while there is a small positive effect of some HE on occupational status for males, the ATE on dropout is both small and not statistically significant.

However, the above is based on a relatively low dimensional setting, and Table 4 presents results based on a higher dimensional model. Here interactions of the baseline variables with ethnicity and female (in the pooled model) are included, along with second and third degree polynomials of the continuous variables. For females, it appears that initial participation in HE is what matters, while for males, graduating and obtaining a qualification is more important. Male dropouts are therefore likely to work in occupations with a lower occupational prestige score than otherwise comparable female dropouts even though, in general, both groups are better off than had they directly entered the workforce upon completion of secondary schooling.

The results in both the low- and high-dimensional setting are stable to tuning parameter choices (see Appendix B for an empirical demonstration), but estimates in the high-dimensional setup benefit from improved precision. The outcomes, however, are similar to those in the low-dimensional setting. For the pooled sample the ATE's are both large and statistically significant,

Table 4: ATE’s Post-LASSO: Dropout and Occupational Status (High Dimensional Setting)

	Pooled		Males		Females	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
ATE						
graduates	0.565327***	0.085439	0.46200***	0.13948	0.6036726***	0.1102535
dropouts	0.435727***	0.101691	0.09511	0.17093	0.6032284***	0.1311772
POM						
graduates	3.7491	0.0424	3.7823	0.0620	3.7520	0.0563
dropouts	3.6195	0.0695	3.4154	0.1157	3.7515	0.0907
EQ only	3.1837	0.0750	3.3203	0.1259	3.1483	0.0962
on/off support	3496/180		1512/44		2029/91	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

though they are slightly smaller than in the low-dimensional setting. In terms of the disaggregated results, when including these additional controls not only is the POM for females who only possess a HE entrance qualification slightly higher, but the gap between males and females shrinks, though males still have a much higher POM on occupational status. For females, completing a HE qualification has almost exactly the same ATE on occupational status as entering HE and subsequently dropping out. Therefore, female dropouts may benefit from their partial HE experience relative to individuals with no HE experience, given they are able to access job roles with a higher occupational status.

This is consistent with the findings of Davies & Elias (2003) who demonstrate that, while dropouts are less likely to be employed than graduates, those who do find employment often find it in a field similar to their field of study, and work in occupations comparable to their graduate peers¹⁰. In the case of the females represented in this sample, not only does the evidence support the findings of Matković & Kogan (2014), but the results are interesting from a theoretical perspective; given they support the idea that entering HE conveys a positive signal. This is not necessarily true in the case of males, however, for whom some tertiary education only slightly

¹⁰These results are unlikely to be driven by those who eventually graduate while on-the-job. So-called "degree apprenticeships" that allow students to gain a bachelor degree via a vocational pathway were only introduced in 2015 (Universities United Kingdom, 2017).

improves occupational status. The results here indicate that, in terms of occupational prestige, female dropouts have relatively better early labour market outcomes than male dropouts.

6 Conclusions

Human capital theory suggests that, if an individual gains skills while (briefly) attending university, they may be more productive than those who only possess a HE entry qualification and so command higher wages. For females the results here indicate no such wage premium; and the hourly wages for dropouts are almost indistinguishable from their peers who possess only a HE entry qualification. On the other hand, signalling theory suggests that enrolling in HE may convey a positive signal to employers even in the absence of a successfully completed course of study. The findings on occupational status are consistent with this idea, and the ATE's for female dropouts and graduates are statistically indistinguishable. For females the evidence is thus more consistent with signalling theory than human capital theory.

For males the opposite is true. Though the ATE of dropout on occupational status indicates a positive effect of some HE experience on occupational status, it is not statistically significant and is small in magnitude - particularly when compared to the large effect size on graduates. For males there is no evidence here to suggest a positive signal is conveyed by partial HE attendance. Conversely, the effect of dropout on wages suggests that male dropouts do receive a wage premium relative to their peers who only possess a HE entrance qualification. The evidence for males is therefore more consistent with human capital theory. In general, however, there is no evidence of a negative penalty associated with dropout for either hourly wages or occupational status. At best, dropouts are better off on one dimension and, at worst, their outcomes are indistinguishable from their peers who never participated in HE. This is not the case when it comes to employment, however, and both male and female dropouts are less likely to be employed than either graduates or those with only a HE entrance qualification.

In terms of potential limitations, omitted variable bias may still remain if the polynomials and interaction terms supplied to the estimator are insufficient for estimating any non-linearities in the conditional expectation function of: a) the likelihood of participating in HE, b) the likelihood of dropping out, and c) the outcomes of interest. However, the model is not sensitive to tuning parameter choices (see Appendix B.1), and there should be no reason to believe that this is the case. Similarly, while positive selection into HE (and subsequent graduation) may upwardly bias differences in wages and occupational status, I find that the results are robust to the inclusion of skill measures in English, Science, Mathematics and ICT measured at age 13/14 (see Appendix

B.2). The results are also consistent in sign with those estimated via OLS (see Appendix C, Table C1), though OLS largely under-predicts treatment effects on wages given dimensionality issues restrict estimation to the low-dimensional setting. OLS estimates of average treatment effects on occupational status, however, are very similar to the DML setting. This is not unexpected, given the DML results on occupational status were similar in the low- and high dimensional settings, and recalling that when the penalty term in the post-LASSO is equal 0 it collapses to the OLS estimator.

What should be kept in mind is that these are early career returns. The time dimension is likely important, as students who successfully completed a tertiary qualification were only a few years post-graduation at the time of the follow-up survey wave. Conversely, if dropouts left HE within or shortly after the first year of study, they may have only slightly less labour market experience than their non-graduate peers. Dropouts may benefit from a signalling or productivity gain, while not being penalised in terms of experience foregone. In terms of avenues for future research, evidence shows this experience penalty for graduates lessens over the course of an individual's working life (see Blundell et al., 2000). It would therefore be an informative exercise to examine the relative trajectories of these individuals. A follow-up wave 9 of the survey is due to be released in 2021, which would allow for the analysis of more medium-term labour market outcomes.

It may also be informative to examine whether the differences in wages found here and in earlier works (i.e. Johnes & Taylor, 1991) are, at least in part, due to changes in cohort composition. Students entering HE after the sweeping changes to the financing of HE likely faced a different incentive structure, both when making the study decision and when making the dropout decision. Not only is the opportunity cost of the initial study decision higher, but the dropout decision is also associated with a higher opportunity cost if students have already invested one or more years worth of tuition payments. Linking the Next Steps survey and the earlier 1970 British Cohort Study to Higher Education Statistics Agency (HESA) data would permit such a cohort comparison, while controlling for the quality of a student's HE entrance qualification. In doing so it may be possible to shed light on such compositional changes.

References

- Aina, C., Baici, E., Casalone, G. & Pastore, F. (2019). Delayed Graduation and University Dropout: A Review of Theoretical Approaches. IZA Discussion Paper No. 12601.
- Angrist, J. & Krueger, A. (1991). Does Compulsory School Attendance Affect Schooling and Earnings. *Quarterly Journal of Economics*, 106(4): 979–1014.
- Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics*, 2(3): 193–216.
- Athey, S., Imbens, G. W., & Wager, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4): 597–632.
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, P2): 9–49.
- Belloni, A., Chernozhukov, V., Fernández-Val, I., & Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, 85(1): 233–298.
- Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2): 521–547.
- Blundell, R., Dearden, L., Goodman, A., & Reed, H. (2000). The Returns to Higher Education in Britain: Evidence from a British Cohort. *The Economic Journal*, 110(461): 82-99.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.
- Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In "Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp," L. Christofides, E. Grant, and R. Swidinsky, (eds.). University of Toronto Press: Toronto, 201–222.
- Carneiro, P., Heckman, J.J., & Vytlacil, E. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6): 2754-2781.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/Debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1): C1–C68.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/Debiased/Neyman machine learning of treatment effects. *American Economic Review Papers and Proceedings*, 107(5): 261–265.
- Chevalier, A. & Conlon, G. (2003), Does it pay to attend a prestigious university?. IZA Discussion Paper no. 848, IZA, Bonn.

- Collins, R. (1971). Functional and conflict theories of educational stratification. *American Sociological Review*, 36: 1002–1019.
- Davies, R., & Elias, P. (2003). *Dropping out: A study of early leavers from higher education*. London: Department for Education and Skills.
- Dearden, L. (1999). Qualifications and earnings in Britain: How reliable are conventional OLS estimates of the returns to education?. IFS Working Paper no. 99/7, The Institute for Fiscal Studies, London.
- Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, 189(1): 1–23.
- Giani, M. S., Attewell, P. & Walling, D. (2020). The Value of an Incomplete Degree: Heterogeneity in the Labor Market Benefits of College Non-Completion. *The Journal of Higher Education*, 91(4): 514–539.
- Gesthuizen, M., de Graaf, P. M. & Kraaykamp G.(2005). The changing family background of the low-educated in the Netherlands: socio-economic, cultural, and socio-demographic resources. *European Sociological Review*, 21(5): 441–452.
- Groot, W. & Oosterbeek, H. (1994). Earnings Effects of Different Components of Schooling: Human Capital versus Screening. *The Review of Economics and Statistics*. MIT Press, 76(2): 317–321.
- Hällsten, M. (2017). Is education a risky investment? The scarring effect of university dropout in Sweden. *European Sociological Review*, 33(2): 169–181.
- Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2): 315–331.
- Harmon, C. & Walker, I. (1995). Estimates of the Economic Return to Schooling for the United Kingdom. *American Economic Review*, 85(5): 1278–1286.
- Heckman J. J., Humphries, J. E. & Veramendi, G. (2016). Returns to education: The causal effects of education on earnings, health, and smoking. National Bureau of Economic Research Working Paper No. 22291.
- Heigle, J., & Pfeiffer, F. (2019a). An Analysis of Selected Labor Market Outcomes of College Dropouts in Germany – A Machine Learning Estimation Approach. Research Report, BMBF, Mannheim.
- Heigle, J., & Pfeiffer, F. (2019b). Langfristige Wirkungeneines nicht abgeschlossenen Studiums auf individuelle Arbeitsmarktergebnisse und die allgemeine Lebenszufriedenheit. *Qualität in der Wissenschaft*, 13(3&4): 95-101.
- Joint Council for Qualifications (JCQ). (2019). GCE A Level GCE AS Level Results Summer

- 2019.
- Johnes, J. Taylor, J. (1991). Non-completion of a degree course and its effect on the subsequent experience of non-completers in the labour market. *Studies in Higher Education*, 16(1): 73–81.
- Knaus, M. (2018). A Double Machine Learning Approach to Estimate the Effects of Musical Practice on Student’s Skills. arXiv preprint arXiv:1805.10300.
- Lechner, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In M. Lechner & F. Pfeiffer (eds.), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica, 2001, 43—58.
- Matković, T. Kogan, I. (2014). Relative worth of a bachelor degree: Patterns of labour market integration among drop-outs and graduates in sequential and integrated tertiary education systems. *Acta Sociologica*, 57(2): 101–118.
- Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.
- Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. *Probability and Statistics*, 57: 213–234.
- O’Leary, N. C. & Sloane, P. J. (2005) The return to a university education in Great Britain, *National Institute Economic Review*, 193(1): 75–89.
- Oreopoulos, P.(2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, 91(11-12): 2213–2229.
- Reisel, L. (2013). Is more always better? Early career returns to education in the United States and Norway. *Research in Social Stratification and Mobility*, 31: 49–68.
- Rosenbaum, P. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70: 41–55.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5): 688–701.
- Schnepf, S. V. (2015). University dropouts and labor market success. *IZA World of Labor*, doi: 10.15185/izawol.182.
- Solomon, L. C. (1975). The Definition of College Quality. *Explorations in Economic Research* 2: 537-87.
- Spence, M.(1973). Job market signaling. *Quarterly Journal of Economics*, 87(3): 355–374.
- Stiglitz, J. E. (1975). The theory of “screening”, education, and the distribution of income. *The American Economic Review*, 65(3): 283–300.

- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58: 267–288.
- UIS (2020). UNESCO Institute for Statistics. data.uis.unesco.org/ (accessed on 20 October 2020).
- University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2018). *Next Steps: Sweeps 1-8, 2004-2016*. [data collection]. 14th Edition. UK Data Service. SN: 5545, doi.org/10.5255/UKDA-SN-5545-6.
- Universities United Kingdom (2017). *Degree Apprenticeships: Realising Opportunities* [Report]. universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2017/degree-apprenticeships-realising-opportunities.pdf.
- Wise, D. (1975). Academic Achievement and Job Performance. *American Economic Review* 65: 350-66.

Appendix A

Table A1: DML: Description & Coding of Control Variables

	coding	# variables generated
Individual Characteristics		
female	binary	1
Family Background		
ethnic background	categorical	6
migrant background (not born in UK)	binary	1
language(s) other than English spoken	binary	1
# siblings by age 17-18	continuous	1
non-single parent household	binary	1
Socioeconomic Characteristics		
social housing tenants	binary	1
welfare (one or more parents)	binary	1
2004 IDACI Score (rounded)	continuous	1
highest parental qualification	categorical	7
Childhood Health		
mothers age at time of birth	categorical	6
birth-weight (kilos)	continuous	1
if premature, # weeks	continuous	1
recorded SEN	binary	1
Cognitive Skills		
frequency reads for pleasure	categorical	6
Own Attitudes (aged 17-18)		
work: job or career is important	categorical	5
attitude to school	continuous	1
School-level Controls		
independent school	binary	1
urban vs. rural	categorical	8
Adulthood Controls		
number of people in household	continuous	1
still living with parent(s)	binary	1

married (or civil partnership)	binary	1
BMI	continuous	1
disability	binary	1
general health score (GHQ12) (Goldberg & Williams, 1988)	continuous	1
locus of control scale (Lefcourt, 1991)	continuous	1
Job-related Controls		
occupational category: NS-SEC 5 (analytic classes)	categorical	6
shift worker	binary	1
spells of unemployment	continuous	1
number of variables		66

Table A2: Distribution of Occupational Status

	EQ-Only		Dropouts		Graduates		Total
	n	%	n	%	n	%	
higher managerial/ professional	345	14.18	519	21.33	1,569	64.49	2,433
intermediate occupations	200	26.39	152	20.05	406	53.56	758
small employers/own accounts	46	40.35	22	19.30	46	40.35	114
lower supervisory/technical	100	46.95	29	13.62	84	39.44	213
(semi-) routine	165	37.5	81	18.41	194	44.09	440
Total	856		803		2,299		3,958

$\chi^2(10) = 265.5755, Pr = 0.000$

Table A3: Average Hourly Wages by Treatment Group

	Total	Males	Females
graduates	8.42	7.91	8.81
dropouts	8.56	8.05	8.88
EQ only	8.87	7.77	9.75

Table A4: ATE's Post-LASSO: Dropout and Hourly Wages (Low Dimensional Setting)

	Pooled		Males		Females	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
ATE						
graduates	-0.36930	0.20876	0.24301	0.17799	-0.56244*	0.28021
dropouts	-0.17349	0.22759	0.40210	0.24475	-0.26497	0.30625
POM						
graduates	8.4929	0.0640	8.0819	0.0895	8.8755	0.0933
dropouts	8.6887	0.1104	8.2410	0.1905	9.1730	0.1554
EQ only	8.8622	0.2005	7.8389	0.1572	9.4380	0.2656
on/off support	2920/40		1185/84		1578/113	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Table A5: ATE's Post-LASSO: Dropout and Occupational Status (Low Dimensional Setting)

	Pooled		Males		Females	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
ATE						
graduates	0.587327***	0.084837	0.55432***	0.12561	0.666638***	0.112358
dropouts	0.477272***	0.100592	0.16432	0.16135	0.692280***	0.131997
POM						
graduates	3.7534	0.0412	3.7894	0.0627	3.7363	0.0574
dropouts	3.6433	0.0679	3.3994	0.1184	3.7620	0.0903
EQ only	3.1660	0.0748	3.2351	0.1100	3.0697	0.0976
on/off support	3632/44		1468/88		2000/120	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Appendix B

B.1 Testing Sensitivity to Tuning Parameters (Knaus, 2018)

One of the extensions to Farrell (2015) proposed by Knaus (2018) is a method for systematically addressing potential sensitivity of the model to tuning parameter choices. This extension is based on the One Standard Error (1SE) rule proposed by Breiman et al. (1984). Fundamentally, sensitivity tests of this kind allow the researcher to assess differences in model performance when using data-driven cross-validated penalty terms (in this case, the penalty term minimising the mean square error (MSE)) instead of theoretical ones. Here I briefly summarise the practical advantages of using such an approach, and their relevance to my empirical setting. For a more detailed theoretical description of the 1SE-rule rule, refer to the original article by Breiman et al. (ibid.).

The extension proposed by Knaus (2018) is based on observation of the following; in the region around the penalty value indicating the *minimum* cross-validated MSE, values of the cross-validated MSE are similar. Thus, there is clearly a degree of uncertainty about the penalty term which minimises the MSE given there may exist a number of plausible values. Unfortunately, depending on which value of the penalty term is ultimately selected, the complexity of resulting model may vary. That is, the model specification may be sensitive to these plausible alternative penalty terms based on the MSE threshold. The original 1SE-rule rule proposed by Breiman et al. (1984) works as follows; first, estimate the standard error (SE) of the cross-validated MSE, then, select the penalty term one SE smaller¹¹. The reasoning behind applying this 1SE-rule is that, if uncertainty about the optimal penalty choice exists, researchers should generally opt for the less complex model.

While based on the 1SE-rule proposed by Breiman et al. (1984), the extension to Farrell (2015) that Knaus (2018) suggests is a little more comprehensive; the idea is to consider a set of models within a range plus or minus one of the estimated SE. Thus, the researcher considers not only a *less* complex model, but also a *more* complex model. Per Knaus (ibid.), this extension is particularly helpful in the DML setting where there are multiple nuisance parameters estimated via separate estimators (i.e. the propensity score and the outcome variable of interest). Recall that there is a two-step estimation procedure used here, the first estimated via OLS, and the second by logistic LASSO. However, the levels of penalties for each of these are not necessarily comparable. This implies that sensitivity checks based on more traditional methods used for ML

¹¹Knaus (2018) notes that, while the choice of exactly one SE is ad-hoc, the 1SE-rule is in fact widely applied in other ML settings.

rather than DML may be inappropriate, as changing the penalty term by either a fixed or ratio based amount might have a very different effect on the nuisance parameter estimated by OLS and the one estimated by logistic LASSO.

This alternative specification instead provides a data-driven way to investigate sensitivity to tuning parameters; one that affects each of the estimators used in the two-step estimation procedure similarly. The following table demonstrates the implementation of this plus or minus 1SE for the empirical setting investigated here, comparing results obtained from estimating -1, -0.5, 0.5 and 1 SE in either direction around the cross-validated minimum. In this case, I focus on the effects on occupational status, and compare the models performance in the low and high-dimensional settings.

Table B1: Occupational Status, 1SE Testing (Basic Model)

	Baseline	SE -1	SE -0.5	SE 0.5	SE 1
ATE (graduates)	0.587563*** (0.084801)	0.604133*** (0.086665)	0.606309*** (0.085305)	0.584810*** (0.086202)	0.581787*** (0.086997)
ATE (dropouts)	0.478720*** (0.101012)	0.496261*** (0.101226)	0.496658*** (0.099761)	0.466851*** (0.101850)	0.463827*** (0.101516)
POM (graduates)	3.7536 (0.0412)	3.7593 (0.0416)	3.7546 (0.0410)	3.7529 (0.0413)	3.7529 (0.0413)
POM (dropouts)	3.6448 (0.0685)	3.6514 (0.0669)	3.6450 (0.0660)	3.6349 (0.0680)	3.6349 (0.0680)
POM (EQ only)	3.1660 (0.0748)	3.1551 (0.0767)	3.1483 (0.755)	3.1681 (0.0764)	3.1711 (0.0773)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Table B1 presents results based on the low-dimensional setting with no interactions or polynomials, whereas Table B2 presents results based on the high-dimensional setting. Results obtained in the low-dimensional setting appear to be slightly more stable than those obtained using the high-dimensional model, however, there are not large differences in either setting. In both cases, moving away from the cross-validated minimum in a negative direction appears to result in slightly larger estimates for the POM's on dropout, and the same is true for graduates in the basic model. The POM on EQ only increases slightly on either side of the minimum in the basic model, but increases slightly when moving in the negative direction in the high-dimensional

Table B2: Occupational Status, 1SE Testing (High Dimensional)

	Baseline	SE -1	SE -0.5	SE 0.5	SE 1
ATE (graduates)	0.566765*** (0.084355)	0.624730*** (0.082576)	0.612620*** (0.084138)	0.574800*** (0.085123)	0.561180*** (0.085945)
ATE (dropouts)	0.422554*** (0.101798)	0.534464*** (0.098320)	0.507376*** (0.099185)	0.404422*** (0.108526)	0.409931*** (0.106734)
POM (graduates)	3.7525 (0.0422)	3.7506 (0.0422)	3.7555 (0.0416)	3.7580 (0.0420)	3.7465 (0.0424)
POM (dropouts)	3.6083 (0.0709)	3.6604 (0.0679)	3.6503 (0.0667)	3.5876 (0.0790)	3.5953 (0.0759)
POM (EQ only)	3.1857 (0.0739)	3.1259 (0.0716)	3.1429 (0.740)	3.1832 (0.0749)	3.18541 (0.0757)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

setting. In general, both models are very stable; the POM for graduates, for example, differs only by 0.0064 between the max and min values in Table B1, and 0.0074 in Table B2.

B.2 Robustness Testing

One potential risk to the CIA is the possibility that, even after selecting a sub-sample of individuals who possess a HE entrance qualification, there is positive selection into HE (and graduation). If this were the case, it may upwardly bias results in terms of overestimating the effect of HE participation on wages and occupational status. In order to test robustness of the results, I repeat the estimation steps using occupational status while including skill measures in English, Science, Mathematics and ICT from tests taken at age 13/14 as a proxy for ability.

Table B3: ATE's Post-LASSO: Occupational Status with Skill Proxies (Low Dimensional Setting)

	Pooled		Males		Females	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
ATE						
graduates	0.508854***	0.084623	0.46110***	0.12702	0.604205***	0.124458
dropouts	0.383618***	0.100412	0.12579	0.15725	0.634032***	0.141464
POM						
graduates	3.7618	0.0414	3.7570	0.0627	3.7585	0.0556
dropouts	3.6365	0.0683	3.4217	0.1111	3.7883	0.0881
EQ only	3.2529	0.0745	3.2959	0.1119	3.1543	0.1121
on/off support	3597/47		1495/44		2048/57	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Table B3 reports results from the low-dimensional setting which, in addition to the function skill measures, consists of the baseline variables from Table A1 with the exclusion of job-related characteristics and the addition of parental occupational status'. Table B4 presents results from the higher-dimensional setting, where interactions of all of the baseline variables (including functional skill measures) with ethnicity and female (in the pooled model) are included, along with second and third degree polynomials of the continuous variables. Comparing the results obtained here with those from Section 5.3, the results are statistically very similar both in terms of effect size and significance level, and the conclusions do not change with the addition of these skill measures.

Table B4: ATE's Post-LASSO: Occupational Status with Skill Proxies (High Dimensional Setting)

	Pooled		Males		Females	
	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
ATE						
graduates	0.498689***	0.086452	0.295592*	0.140738	0.535175***	0.110879
dropouts	0.379072***	0.102819	0.013187	0.0173095	0.570842***	0.130007
POM						
graduates	3.7504	0.0421	3.7481	0.0636	3.7497	0.0579
dropouts	3.6227	0.0697	3.4657	0.1178	3.7854	0.0900
EQ only	3.2517	0.0764	3.4525	0.1267	3.2145	0.0958
on/off support	3541/103		1481/58		1950/155	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

Appendix C

Table C1: OLS Estimation of Hourly Wage Effects: Low-Dimensional Model

	Coef.
Treatment	
<i>(ref. cat. HE EQ)</i>	
Dropout	0.0196 (0.184)
Graduate	-0.000558 (0.160)
Female	1.191*** (0.114)
Ethnicity	
<i>(ref. cat. White)</i>	
Mixed	-2.238*** (0.276)
Indian	-1.902*** (0.230)
Pakistani & Bangladeshi	-2.092*** (0.291)
Black	-1.861*** (0.279)
Other	-2.365*** (0.412)
Migrant Background	0.206 (0.267)
Language(s) other than English Spoken	0.290 (0.236)
# of Siblings	-0.00907 (0.0556)
Non Single Parent Household	0.712*** (0.154)
Social Housing Tenant	-0.128

	(0.207)
Welfare (one or more parents)	-0.263**
	(0.134)
IDACI Score	-0.409
	(0.424)
Highest Parental Qualification	
<i>(ref. cat. No Qualifications)</i>	
Other Qualifications	0.0782
	(0.408)
Qualifications at Level 1 and Below	0.559*
	(0.289)
GCSE Grades A-C or Equivalent	1.102***
	(0.207)
GCE A Levels or Equivalent	0.990***
	(0.228)
HE Below Degree Level	1.075***
	(0.226)
Degree or Equivalent	1.046***
	(0.226)
Mother's Age at Birth	
<i>(ref. cat. Under 20)</i>	
20-24	0.0592
	(0.326)
25-29	-0.0189
	(0.320)
30-34	-0.120
	(0.327)
35+	0.174
	(0.354)
Unknown	0.486
	(0.856)
Birth-Weight (KG)	0.115
	(0.115)

Weeks Premature	0.0330 (0.0368)
Frequency Reads for Pleasure (<i>ref. cat. Unsure</i>)	
Never	1.562 (1.556)
Hardly Ever	2.116 (1.536)
Less than Once a Week	1.955 (1.541)
Once a Week	2.170 (1.537)
More than Once a Week	1.858 (1.541)
Most Days	1.952 (1.538)
Job or Career is Important (<i>ref. cat. Strongly Agree</i>)	
Agree	0.712*** (0.156)
Disagree	0.494 (0.600)
Strongly Disagree	-0.210 (0.999)
Attitude to School	-0.0234* (0.0135)
Urban vs. Rural (<i>ref. cat. Urban, Sparsely Populated</i>)	
Town & Suburbs, Sparsely Populated	0.310 (2.262)
Village, Sparsely Populated	-1.390 (2.227)
Hamlet or Isolated, Sparsely Populated	-0.740

	(2.218)
Urban, Densely Populated	-0.288
	(2.107)
Town and Suburbs, Densely Populated	-0.169
	(2.114)
Village, Sparsely Populated	-0.0800
	(2.115)
Hamlet or Isolated, Less Sparse	-0.365
	(2.129)
Independent School	-0.424*
	(0.235)
# of People in Household	0.205***
	(0.0463)
Still Living with Parent(s)	-0.423***
	(0.154)
Married or in Civil Partnership	0.317*
	(0.182)
BMI	-0.0227***
	(0.00824)
General Health Score	-0.0128
	(0.0200)
Disability	0.232
	(0.211)
Locus of Control Scale (Lefcourt, 1991)	0.0362
	(0.0340)
Occupational Category	
<i>(ref. cat. DK)</i>	
Semi-Routine & Routine Occupations	1.198
	(0.847)
Lower Supervisory and Technical	-0.494
	(0.867)
Small Employers & Own Accounts	0.353
	(0.895)

Intermediate Occupations	-0.383
	(0.838)
Higher Managerial & Professional	-0.906
	(0.833)
Shift Worker	0.0644
	(0.135)
Unemployment Spells	0.114**
	(0.0577)
Constant	5.529**
	(2.746)
<hr/>	
Observations	3,048
R-squared	0.210
<hr/>	
Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$	
<hr/>	

Table C2: OLS Estimation of Occupational Status: Low-Dimensional Model

	Coef.
Treatment	
<i>(ref. cat. HE EQ)</i>	
Dropout	0.559*** (0.100)
Graduate	0.660*** (0.0868)
Female	0.154** (0.0635)
Ethnicity	
<i>(ref. cat. White)</i>	
Mixed	-0.161 (0.149)
Indian	0.239* (0.129)
Pakistani & Bangladeshi	0.195 (0.161)
Black	-0.119 (0.151)
Other	-0.0445 (0.219)
Migrant Background	-0.108 (0.143)
Language(s) other than English Spoken	-0.0480 (0.129)
# of Siblings	-0.0139 (0.0310)
Non Single Parent Household	0.255** (0.107)
Social Housing Tenant	-0.259** (0.113)

Welfare (one or more parents)	-0.119 (0.0759)
IDACI Score	-0.115 (0.239)
Highest Parental Qualification <i>(ref. cat. No Qualifications)</i>	
Other Qualifications	0.0821 (0.220)
Qualifications at Level 1 and Below	-0.276* (0.159)
GCSE Grades A-C or Equivalent	0.0271 (0.116)
GCE A Levels or Equivalent	0.0563 (0.128)
HE Below Degree Level	0.0275 (0.128)
Degree or Equivalent	0.0509 (0.128)
Mother's Age at Birth <i>(ref. cat. Under 20)</i>	
20-24	0.304* (0.179)
25-29	0.275 (0.176)
30-34	0.274 (0.180)
35+	0.0621 (0.193)
Unknown	0.923** (0.465)
Birth-Weight (KG)	0.108* (0.0641)
Weeks Premature	0.00602

	(0.0209)
Frequency Reads for Pleasure (<i>ref. cat. Unsure</i>)	
Never	-2.789*** (0.906)
Hardly Ever	-2.578*** (0.896)
Less than Once a Week	-2.546*** (0.899)
Once a Week	-2.661*** (0.897)
More than Once a Week	-2.663*** (0.899)
Most Days	-2.747*** (0.897)
Job or Career is Important (<i>ref. cat. Strongly Agree</i>)	
Agree	-0.212** (0.0866)
Disagree	0.0465 (0.333)
Strongly Disagree	-0.436 (0.512)
Attitude to School	0.0293*** (0.00783)
Urban vs. Rural (<i>ref. cat. Urban, Sparsely Populated</i>)	
Town & Suburbs, Sparsely Populated	1.187 (1.008)
Village, Sparsely Populated	1.793* (1.008)
Hamlet or Isolated, Sparsely Populated	1.316 (1.005)

Urban, Densely Populated	1.587*
	(0.922)
Town and Suburbs, Densely Populated	1.731*
	(0.928)
Village, Sparsely Populated	1.655*
	(0.929)
Hamlet or Isolated, Less Sparse	1.689*
	(0.938)
Independent School	0.353***
	(0.136)
# of People in Household	-0.0473*
	(0.0254)
Still Living with Parent(s)	-0.0559
	(0.0855)
Married or in Civil Partnership	-0.122
	(0.102)
BMI	-0.00275
	(0.00454)
General Health Score	-0.0454***
	(0.0108)
Disability	-0.533***
	(0.109)
Locus of Control Scale (Lefcourt, 1991)	-0.0852***
	(0.0188)
Unemployment Spells	-0.0866***
	(0.0312)
Mother's Occupational Status	
<i>(ref. cat. DK)</i>	
Unemployed	-0.117
	(0.123)
Semi-Routine & Routine Occupations	0.0549
	(0.127)
Lower Supervisory and Technical	0.116

	(0.184)
Small Employers & Own Accounts	0.170
	(0.181)
Intermediate Occupations	0.0876
	(0.113)
Higher Managerial & Professional	0.0264
	(0.174)
Father's Occupational Status	
<i>(ref. cat. DK)</i>	
Unemployed	0.108
	(0.137)
Semi-Routine & Routine Occupations	-0.174
	(0.132)
Lower Supervisory and Technical	-0.252*
	(0.141)
Small Employers & Own Accounts	-0.109
	(0.125)
Intermediate Occupations	0.0657
	(0.110)
Higher Managerial & Professional	-0.0616
	(0.125)
Constant	4.212***
	(1.272)
<hr/>	
Observations	3,676
R-squared	0.097
<hr/>	
Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$	