What effect does School Absence have on an individual's Labour Market Outcomes?

Tobin Lochrie
Professional Economist BSc and Apprenticeship Level 6
School of Economics
University of Kent, July 2023

Abstract

When pupils are absent from school, they miss out on valuable teacher-led instruction, which is fundamental to developing their human capital (skills and knowledge). As a result, an individual's productive capabilities are decreased, which could lead to diminished labour market outcomes. It is therefore concerning that attendance has not returned to the levels seen prior to the Covid-19 pandemic school closures. This paper finds: School absenteeism is associated with a reduction in total annual earnings. Additionally, school absence is linked to an increased risk of claiming benefits for at least 6 months, reduced odds of being employed, and increased likelihood of being on a low income (>15K p.a.) at age 28.

Acknowledgements

I want to express my gratitude to Oliver Anderson for his guidance and insights, which have been instrumental in ensuring the highest standards were in place for this paper. I would also like to thank Moira Nelson for her help all the data requests.

1. Introduction

The negative shock of the Covid-19 pandemic, and the disruption caused by the resulting school closures has had a scarring effect on attendance in English schools. Published Department for Education (**DFE**) statistics, shows *overall absence* was consistently below 5% between Autumn term 2016/17 and 2019/20 (the last unaffected by the Covid-19 pandemic). Since Autumn term 2019/20, when the rate was **4.9%,** overall absence rates have risen sharply (Figure 1). Autumn term 2021/22 was 2.0pp higher than 2019/20 at **6.9%** and rose an additional 1.0pp to **7.9%** the following Autumn term (2022/23), amounting to more than ~33m missed days during this period (Explore Education Statistics, 2023).

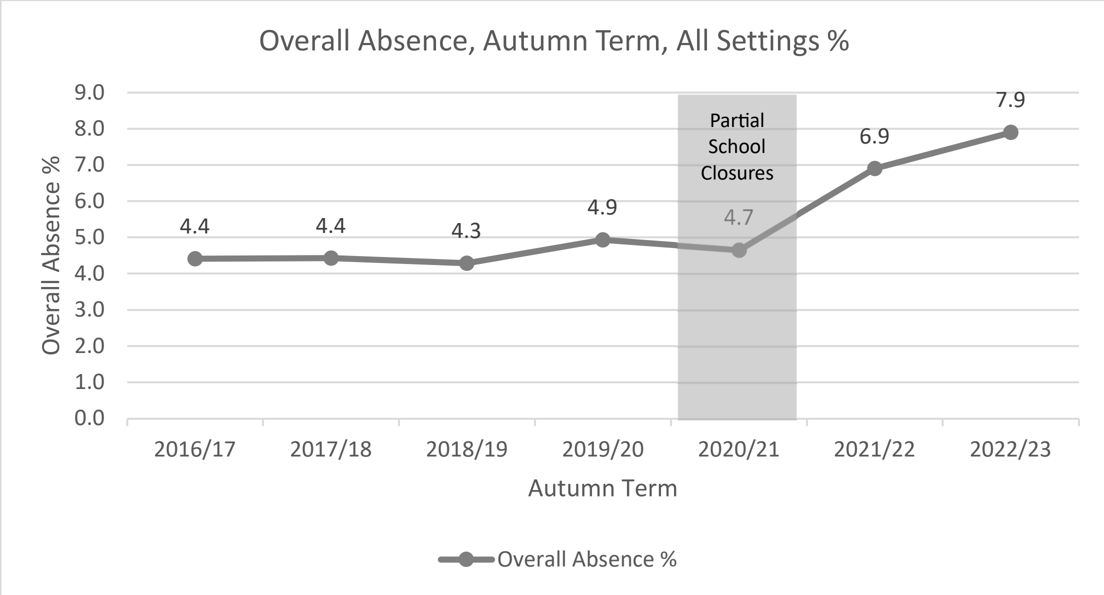

Figure 1 (Autumn 2020/21 rates were affected partial school closures) Source DFE 2023

Moreover, the percentage of enrolments who miss 10% or more of possible sessions, (*the DFE-defined "persistent absent" (**PA**) rate*), is also higher post pandemic. Figure 2 shows the steep rise in PA since the pandemic. Between the 2019/20 and 2022/23 Autumn terms, the PA rate increased by **~12.0pp** from **13.1%** to **25.0%. ~1.6m** pupils missed at least 7 days of school throughout Autumn term 2021/22, compared to **~923k** in Autumn 19/20 (Explore Education Statistics, 2023).

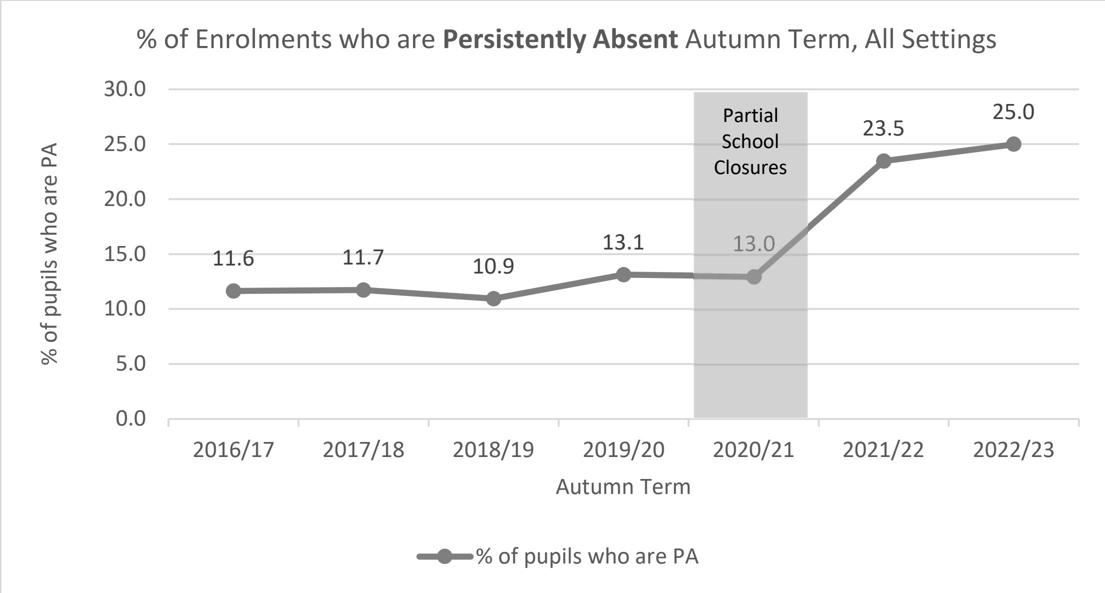

Figure 2 Persistent Absence Rates 2016/17-22/23 Source DFE 2023

The struggles experienced during the lockdowns and extended disengagement from school are cited as potential reasons for the increase in persistent absence (Dräger et al, 2022). Other anecdotal evidence points to issues such as anxiety or depression, coupled with a shortage of mental health provision as a driving force (Fortin 2022).

These figures are concerning, as evidence suggests absence has a direct negative impact on school attainment (DFE (2016), Sims (2020), Gottfried (2011), and Aucejo & Romero (2016)). Additionally, it widely believed that absenteeism will also affect an individual's career prospects, but this effect is a relatively underdeveloped area of research. This paper aims to fill this knowledge gap by answering the question:

"*What effect does school absence have on an individual's labour market outcomes?".*

A range of econometric techniques will the deployed to analyse the private returns to attendance, using a previously unavailable dataset which combines individual-level data on school absence and earnings, with socio-demographic characteristic data.

Analysis focusses on absence during Key Stage 4 (**KS4**) and the effect it has on medium-term labour market outcomes by:

i. Establishing the odds of absent pupils being *employed*, claiming *benefits* at age 28 or having a *low income* (<£15k p.a.).
ii. Modelling the *marginal effect* of a 1pp increase in *absence on earnings* at age 28.

Answering these questions will enable policy makers to make better informed decisions, as the opportunity cost of absenteeism can be monetised in cost benefit analysis models. More significantly, a tangible figure on lost earnings - as a result of absence, can be used to disincentivise poor attendance. Previous *behavioural economics* research suggests *loss aversion* (in this case potential lost income) has a greater influence on individual's behaviour than implied future gains (Tversky and Kahneman 1991).

2.1 Economic Theories of Human Capital and Education

The economist *Gary Becker* (1964) posited that, the unobserved variation in output growth, after controlling for physical capital and labour force accumulation, could be attributed to factors such as the skill, training, and experience of the workforce (collectively referred to as **human-capita**l). At a *Macro* level, a well-trained and highly skilled workforce could lead to higher output and greater economic growth.

Several papers investigate the private returns to education. A recent paper by Hodge et al. (2021) found that a one-grade improvement in General Certificate of Secondary Education (GCSE) attainment, resulted in an *£8,500* (present-value, 2021 discounted prices) increase in lifetime earnings. A standard deviation (11.2 grades) improvement would increase lifetime earnings by *£96,000*. The authors acknowledge there is variation within this broad value; the marginal returns of improving from a D grade to a C grade in Mathematics, for a non-disadvantaged white male, are relatively higher than lower grade improvements for other socio-demographic groups. They also note that their results cannot determine if the improvements are due to increasing a student's skills and knowledge (*human-capital theory*) or the returns are caused by the possession of a certificate which *signals* the learner's skills, motivation, and work ethic to an employer, irrespective of whether the grade improvement has increased their ability (*signalling theory*).

Hodge et al's analysis uses the **Longitudinal Educational Outcomes** (***LEO)*** dataset, which links individual pupil data from the **Department for Education** to earning data from **HMRC** tax records and benefits records from the **Department for Work and Pensions**. There are also a number of studies which have attempted to monetise returns to GCSEs using the **Labour force survey** (**LFS**) (Greenwood et all (2007) and Mackintosh (2006)). They estimated that obtaining 5 or more "*good*" GCSEs had marginal returns of between ~25% to ~31% compared to those who hold no qualifications. While there are some demographic characteristics in the LFS, it lacks disadvantage (FSM) and Special Education Needs and Disabilities (SEN) data (which are within the LEO dataset) which may have explanatory power on future earnings. In addition, the LFS survey relies on individuals reporting their earnings honestly and could therefore lack the accuracy and robustness of the LEO.

At the *Micro* level, individuals can improve their human-capital and productive capabilities by investing their time in education and training. This could potentially lead to greater private wage returns (Hodge et al. 2021). Conversely, high absence might result in individuals lacking the human-capital skills required for positive outcomes in the labour market (Heckman, 2006). Increasing human-capital stocks would, in theory, benefit individuals, firms, and the wider economy.

2.2 Literature Review

There is a wealth of literature that attempts to estimate the impact of absence on attainment. However, due to data availability, there are fewer studies on the effect it has on labour market outcomes. This review critiques the evidence gathered so far.

2.2.1 Current Research Into The Impact On Labour Market Outcomes

There are two notable studies that investigate the result of absenteeism on employment status and earnings. Both papers focus on the longer-term and find non-attendance at school is associated with negative labour market outcomes.

Dräger et al's research (2022) points to a greater likelihood of "non-employment" for poor attenders (as a possible consequence of poor attainment). They found that individuals in their sample who missed 5 days of school, at age 10, were 5.7% more likely to obtain no qualifications and 4.2% more likely to be "not employed" at age 42 than those with 0% absence. However, they were unable to establish a statistically significant relationship between absence and future earnings. Conversely, Cattan et al (2022) found that 10 days absence per year could result in 1%-2% lower lifetime earnings.

Dräger et al used data from the 1970 British Cohort Study (BCS), a panel dataset of 13,776 participants, born during a single week (April 5-11) in 1970. The BCS records in-depth data on educational attainment, school absences during spring term 1980/81, and labour market returns at various points up to mid adulthood (age 40+). Absence was reported manually by the individual's teacher, making it prone to human error.

Cattan's study focusses on Swedish individuals born between 1930-35. Absence data was taken from 1940s school records and joined to corresponding earnings data from population census records in 1962 and 1972. Pension and tax data was used to calculate lifetime earnings. Their study highlights many unobservable confounding variables that could impact school absence (less engaged parents, uninspiring teachers, poor health) which may lead to bias results. To deal with endogeneity in their models, they use within-family (between siblings) fixed effects in their models.

Both Catan and Dräger's studies are limited as they use absence records from late childhood (up to age 10/11). Moreover, Dräger et al's absence data was from a short collection window (January to Easter 1981). Moreover, self-employed individuals are excluded from the earnings regression. Additionally, both datasets may be prone to error given the age of the records and collection methods. All of which may explain why Dräger et al found no statistically significant link between absence and earnings.

The analysis within this paper advances this area of knowledge by addressing these limitations by:

- Considering absence during the GCSE period (age 14-16) which may have a greater impact on attainment and employment.
- Collecting absence data for a longer period (two years)
- Using an up-to-date reliable source of absence and earnings data (LEO)

This will add value to the evidence base on absence and provide policy makers with more robust results.

2.2.2 Research Into The Effects Of School Absence On Attainment.

The consensus amongst academics is that absence has a detrimental impact on school performance. Entwisle et al (2001) likened being in school to "running a faucet", arguing that frequent attendance in school is necessary to build skills and knowledge (human-capital). Once the "faucet is turned off" (i.e. pupils stop attending school) their progress will suffer, which will be reflected in poor attainment. Many other studies (DFE (2016), Sims (2020), Gottfried (2011), and Aucejo & Romero (2016)) agree that absence affects attainment, however the magnitude of the impact varies from paper to paper.

Published analysis by the **Department for Education (2016)** on a sample of ~628.5k pupils within English schools, who took their GCSEs in 2014, found a reduction of around 3.2% in the likelihood of achieving 5 A*-C grades (a "pass") at GCSE, for each day missed (while controlling for similar characteristics). The odds for PA pupils would therefore be reduced by **68%** (10% of a school year = 19 days x 3.2%). The paper also finds that pupils with no absence were 2.2x more likely to achieve 5+ GCSE passes, than pupils missing 10-15% of lessons during KS4. These figures are likely to be reliable as they are based on a large dataset of English students, which contains characteristic data other studies may not have access to.

Meta analysis by Sims (2020) estimates that each day of absence resulted in ~0.3-0.4% of a standard deviation reduction in achievement. Included in his analysis are 2 studies on American pupils aged 6-13. The first, by Gottfried (2011) suggests a marginal day of absence results in a 10%-14% of a standard deviation reduction in achievement (a PA student would therefore see a full standard deviation reduction). In contrast, a study by Aucejo & Romano (2016) which controlled for institutional heterogeneity by including school, teacher, and individual fixed effects, found a smaller reduction in achievement of ~0.3-0.7% of a standard deviation.

The Education Endowment Foundation (2018, pp.26–28) calculated their own figure of 0.53% of a standard deviation by dividing Glass et al's (1981) effect size (+1 for each additional year of schooling) by 190 (the typical number of days in a school year). These effect size estimations are limited as they aren't based on studies conducted in England, and the Glass study is over 40 years old.

Disadvantage is a driver of both high absence and lower labour market outcomes. (Lázaro et al (2020), Beynon and Thomson (2021) and DFE (2018), find that pupils from low-income families are more likely to be absent from school than their peers. Beynon and Thomson (2022) found FSM pupils tend to achieve much lower Reading and Maths scores than their peers with *the same level of absence* (at age 10). Research form the DFE (2018) (using LEO data) found that FSM-eligible pupils were 23% less likely to be in sustained employment aged 27 when compared to non-FSM peers. Due to the initial lower labour market outcomes for FSM pupils, GCSE improvements for this cohort lead to a higher returns. Hodge et al (2021) found FSM eligible pupils see a 9% larger marginal returns to a GCSE grade improvements compared to non-disadvantaged peers.

The Gottfried (2011), Aucejo & Romano (2016), Education Endowment Foundation (2018) have limitations due to location (not based in English schools), varying outcomes, smaller sample sizes and age. Moreover, they do not address how the findings translate into real world outcomes for individuals. The analysis within this paper attempts to address these limitations by:

- Focussing on pupils in English schools
- Including self-employed earners
- Using a recent reliable data source (LEO) with a large sample size (n = 566k)
- Using the findings to assess the economic impact of absence.

3.1 Data

Modelling is conducted using the same **Longitudinal Educational Outcomes** (***LEO)*** panel dataset used in the Hodge et al (2021) and DFE (2016) studies. Since data is obtained directly from government data sources, the analysis will be more robust than any model that relies on survey answers.

The Analysis focusses on a single cohort of pupils who took their GCSEs during academic year 2006/07. Absence data was not available for individuals who went to (private sector) Independent schools. Therefore, model estimates only apply to those who attended (public sector) state-funded schools.

Data was prepared prior to analysis, to ensure only complete records were used. Pupils with "NA" for "FSM", "SEN", "Minor Ethnic Group", "Major Language" and individuals with fewer than 1 potential session were removed. This may affect some results; a pupil who only has 10 potential sessions, would be classed as PA if they missed 1 session. A typical 2-year KS4 period, could have up to ~760 potential sessions, depending on how schools record final-year study leave.

Detailed descriptive statistics for selected subgroups, within the final dataset comprised of **566,181 individuals,** are displayed in Table 1.

	Number of Pupils (n)	**Mean Earnings At 28**	**Mean Overall Absence**	**Mean KS4 Points**
Total	566,181 *(100% of population)*	£24,334.81 p.a. *(Standard Deviation 26764.23)*	9.2% *(Standard Deviation 9.82)*	43.8 *(Standard Deviation 21.3)*
Male	287,046 *(51%)*	£27,279.23 p.a. *(SD 34,175.7)*	8.7% *(SD 9.6)*	41.2 *(SD 21.3)*
Female	279,135 *(49%)*	£21,201.31 p.a. *(SD 14,718.3)*	9.6% *(SD 10.1)*	46.5 *(SD 20.9)*
Not Free School Meals only	519,397 *(91.7%)*	£24,690.16 *(SD 26775.41)*	8.93% *(SD 9.58)*	44.42 *(SD 21.29)*
Free School Meal Only	46,784 *(8.3%)*	£20,012.75 p.a. (SD 26245.84)	12.07% (SD 11.83)	37.28 (SD 19.82)
Not Special Educational Needs only	487,822 *(86.2%)*	£24,980.93 p.a. *(SD 27319.65)*	8.6% *(SD 9.1)*	46.8 *(SD 20.2)*
Special Educational Needs only	78,359 *(13.8%)*	£19,851.40 p.a. *(SD 22020.26)*	13.0% *(SD 13.0)*	25.4 *(SD 18.4)*
Free School Meal & Special Educational Needs	22243 *(3.9%)*	£16467.63 p.a. *(SD 11941.3)*	17.0% *(SD 15.5)*	19.0 *(SD 16.6)*
Persistent Absent (PA)	168,114 (29.7%)	£19,792.46 p.a. *(SD 24065.93)*	19.9% *(SD 12.1)*	30.5 *(SD 20.2)*
Non- Persistent Absent	398,067 (70.3%)	£25,996.03 p.a. *(SD 27498.73)*	4.7 % *(SD 2.62)*	49.4 *(SD 19.1)*

Table 1 Descriptive statistics for key variables.

3.2 Variables

3.2.1 Key Explanatory Variables

Absence Data

"***Total absence rate***" is calculated using the DFE's standard approach: total missed sessions divided by total potential sessions (1 day = 2 sessions). Total sessions are for the whole of KS4 (2 years/11 half terms- data is not collected of the final half term of year 11 as they are on study leave) when pupils are age 14-16 (which is a larger window than Dräger's model, which covers a single spring term – 2 half terms). The age range differs from both Dräger and Cattan's studies (which focus on 10/11-year-olds). There is no implication pupils develop better human capital at this age, only KS4 is a critical period when pupils study for their GCSEs, the results of which will affect future outcomes.

"***Persistently Absent***" **(PA)** is defined as missing more than 10% of possible sessions (in line with the **DFE**'s *current* definition) during KS4, equivalent to between 30-38 days over two years. ~168k (30%) pupils were classed as PA in the LEO dataset. This group accounted for a disproportionate ~65% of all KS4 absence sessions.

Absence data is strongly right skewed, with a mean of 9.2% (28 Days) and a median of 6.3% (19 days) absence over the 2 academic years (Figure 3).

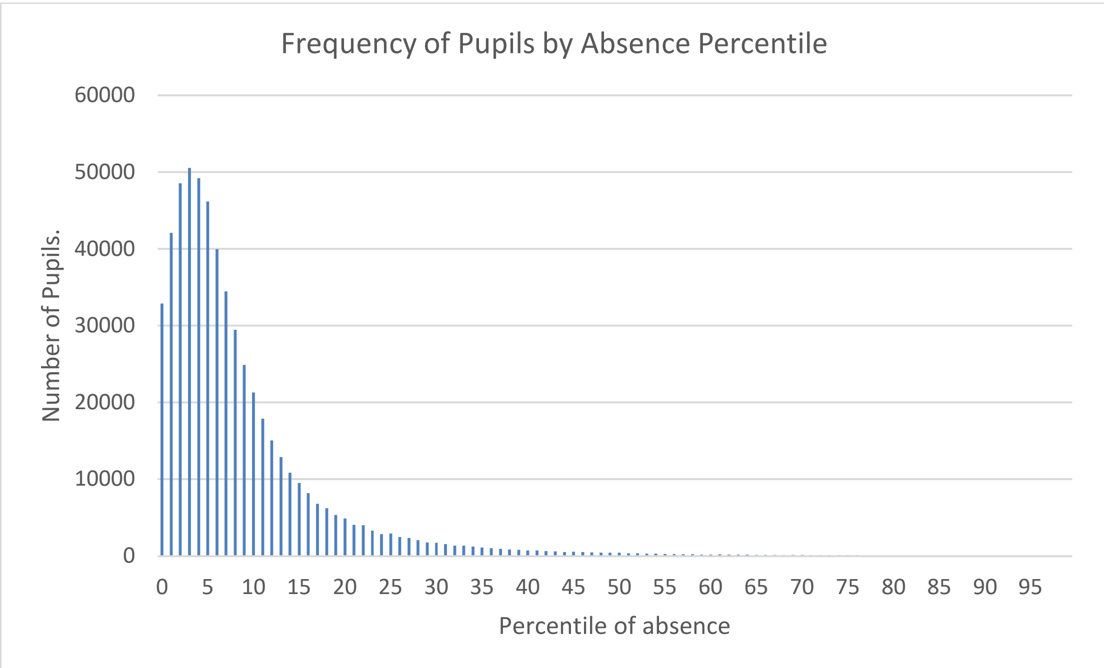

Figure 3 Number of Pupils by percentile of absence Source LEO

- 5.8% of the pupils (32,868) had 0% absence.
- 70% (398,102) missed less than 10% of possible sessions.
- 0.01% (42) missed 100% of possible sessions.
- 30% of pupils (168,170) were **PA** over the two academic years. (This figure appears higher than the published rate for 2006/07 (19.3%) however, published data only covers one academic year and uses a different methodology for preparing data for publication).
- 1.1% of pupils (6,420) missed more than 50% of possible sessions.

3.2.2 Dependent Variables

Total Earnings at 28

"Total Earnings" is an individual's employed (PAYE) or declared self-employed gross yearly earnings (unlike Dräger et al (2022) who excludes these individuals) from tax year 2019/20- when the individuals are age 28. All earnings values within this paper have been inflated to 2021 prices using an ONS GDP deflator. The natural logarithm of "*total earnings*" was taken to deal with its right-skewed distribution. Members of the cohort who are not employed or economically inactive are excluded from earnings modelling.

The tax year 2019/20 was unaffected by the furlough scheme (launched on 20th April 2020 in the following tax year). However, it should be noted that the first UK lockdown commenced on the 23rd of March 2020, 13 days before the tax year ended, which could have a minor impact on earnings. Using earnings data from 2020/21 onwards was rejected due to the potential for the scarring effects of the Covid-19 pandemic and the impact of the 2022/23 "cost-of-living crisis".

Bivariate analysis demonstrates a negative relationship between average earnings and absence percentile (figure 4).

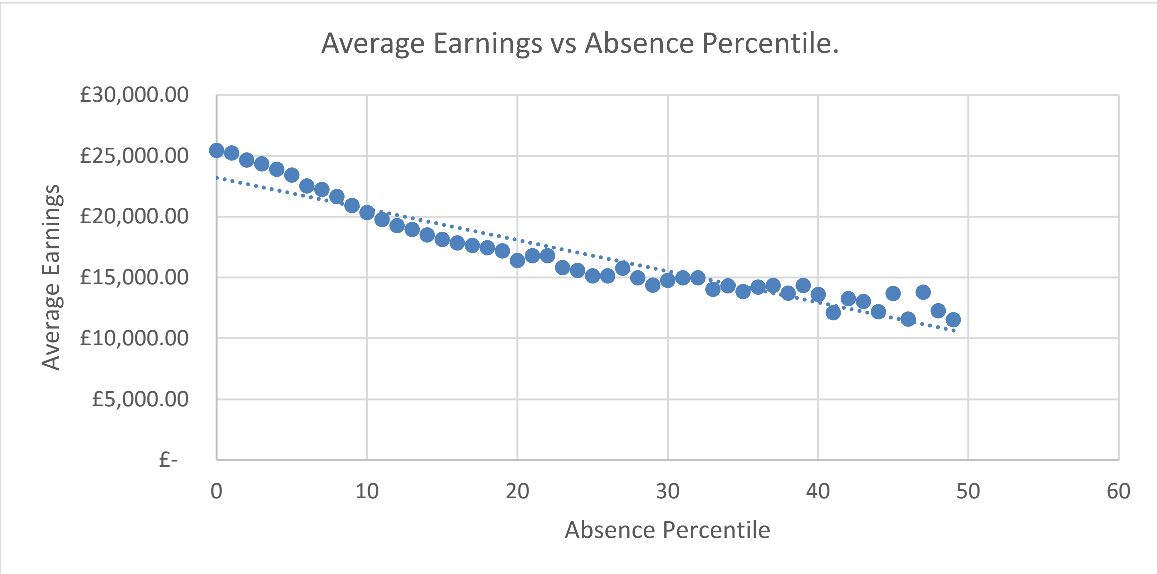

Figure 4 Average Earnings vs Absence Percentile (1.1% of the cohort are absent more than 50%, average earnings are influenced by outliers and therefore not included in this plot). Source LEO

Employed for a Sustained Period

The ~375k Individuals who were employed for at least 1 day each month or had earnings greater than 0 for the whole tax year are classed as "*Employed*" in the modelling. This variable also includes those who are self-employed. It is important to note, those who are not classed as employed (n= ~102k) are not necessarily "unemployed" (not employed *but* actively seeking work) but may also be *economically inactive* (not part of the labour force *and* not seeking employment), which will include those with a disability, caring responsibilities and (mature) students.

Claiming benefits at 28

~60k members of the cohort claimed benefits for at least 1 day in a month for 6 months. These individuals are classed as "**Claiming benefits at 28"**. It should be noted "benefits" are not necessarily unemployment insurance, but also include incapacity benefits, for individuals who are seeking employment but are not well enough to work.

Low Income (earning under £15k p.a.)

As discussed in the limitations, the LEO dataset doesn't record the number of hours worked. It is difficult to establish if an individual works full or part time hours. However, the minimum wage for this cohort would have been £8.21 per hour (GOV.Uk, Minimum Wage 2023) during tax year 19/20. The average working week in February 2020 was 36.9 hours (Statista, 2023) with 52 weeks in a year the minimum salary someone on a full-time contract could earn is roughly £15,753 (8.21 x 36.9 x 52). This variable captures the ~78k individuals who either:

- did not have a full-time job for the whole tax year,
- have unstable hours (zero-hours contracts, "gig economy" etc)
- work part time hours*,

These individuals will be recorded as "*low income*" in the data. Those in the 30th (and higher) absence percentiles, on average fall into this category (figure 4).

*There is uncertainty around this definition as full-time hours differ between employers. Additionally, this variable does not capture those who work part time/unstable-hours but earn over £15k during the tax year.

3.2.2 Additional Earnings-Influencing Controls

Previous work on human capital with the LEO dataset controlled for the following socio-demographic factors that could influence school and employment outcomes: **Gender**, **Major Ethnic Group, Special Educational Needs (at age 16), Free School Meal Eligibility (at age 16) and School.**

Individuals are likely to be clustered into non-random groups (their school, location at 16). There may be similarities in attainment and earnings for those at the same school, for example those who attended *selective* schools have higher average earnings (£36,380.63) and lower absence rates (7.1%) than those in *non-selective* schools (£23,842.76 and 9.3% respectively).

Additionally, there will be school-specific *unobservable* time-invariant factors (leadership, teaching standards, school culture) and *observable* ones (ofsted rating, religious-ethos, location, and selection policy) which may influence an individual's absence achievement, and labour market outcomes. Like Aucejo & Romano (2016) school-level heterogeneity is controlled for by including the school's unique reference as a **fixed effect**.

3.3 Full Table Of Variables

Name	Definition
Total Earnings	Total (Employed/Self Employed) earnings in tax year 2019/20
Claiming benefits at 28	=1 if in receipt of benefits 1 day a month for at least 6 months, 0 otherwise.
Total Absence	Number of sessions missed in year 10 and 11 divided by number of potential sessions.
Persistent Absent	=1 if absent for more than 10% of potential sessions, 0 otherwise
Major Ethnic Group	Categorical variable, Black, Asian, Mixed Race, Other, unclassified, and White (reference group)
Region at 28	Categorical Variable, Region lived in when earning. Reference London.
Free School Meal	= 1 if eligible for Free School Meals at 16 (but *no* Special Educational Need), 0 otherwise
Special Educational Needs	=1 if has a Special Educational Needs diagnosis at 16(but *not* eligible for Free School Meals), 0 otherwise
Free School Meal & Special Educational Needs	= 1 if eligible for Free School Meals *and* has Special Educational Need diagnosis at 16, 0 otherwise
Female	=1 if female, 0 if male
Low Income <£15k p.a.	=1 if earnings >=£15,000 p.a. at 28, 0 otherwise
School Unique Reference	Included as a fixed effect.
Attending University or Higher	=1 if by 28 the individual had at least a level 6 as their highest level of education, 0 otherwise. (*Level 1 -leaving school at 16 with no GCSE passes, to Level 8- doctorate, PhD*).
Key Stage 4 Points	Points: 8 for an A*, 7 for an A…. to 1 for a G.

Table 2

3.4 Limitations

The LEO dataset has limitations which may lead to omitted variable bias and thereby affect the estimators:

Firstly, the scope of this study was restricted by the number of years covered: 28 is the latest age at which earnings could be compared to school attendance. Evidence suggests that median weekly pay peaks in the 40-49 age bracket. (House of Commons Library, 2022). Despite this, it is felt that 28 is a reasonable age to measure the impact of absence, given the majority of individuals will be active in the labour market by this age.

Secondly, it wasn't possible to differentiate types of absence (*e.g. illness, vacation, religious holidays, truancy, etc*). This makes it difficult to distinguish between those who are absent for health reasons (and motivated to catch up missed learning) and those who are truant (whose academic and therefore labour market outcomes may vary). This granularity of absence is recorded for later GCSE cohorts however their earnings would be recorded at a much younger age.

Thirdly, observable variables such as *union membership* aren't included in the models. Card (1996) suggested *union membership* raised the wages of individuals with lower skill levels in a 1996 study. This may still be the case in some sectors however it is not recorded in the LEO dataset. Additionally, labour economist Jacob Mincer argues "*potential work experience"* has an impact on wages (Heckman et al, 2003). Models typically calculate this as: *age, minus years schooling, minus six*. As this study uses a single GCSE cohort and roughly all individuals are aged 28, the value for "potential work experience" would not differ between university graduates in the dataset and would therefore only be suitable in models containing multiple GCSE cohorts. There is a possibility union membership and work experience could be influenced by the key explanatory variable (absence) and therefore their inclusion may result in endogeneity.

Additionally, LEO does not record hours worked. Working part time is a labour market outcome, the reason for which may vary. According to European Union statistics, a significant number of women choose to work part-time in order to care for children or family members (Folguera et al, 2022). Including this variable as a control is problematic as it occurred after (and therefore may be a result of) the school absence. Men often cite the reason for working part time as "*struggling to find full-time employment*" (Folguera et al, 2022). They, and women without childcare responsibilities, may struggle to obtain full-time employment due to poorer academic performance as a consequence of poor attendance.

Finally, (similar to the Cattan Study 2020) there are unobservable variables such as individual preferences, innate ability, personality, parental engagement, and health status which are difficult to record and not included in the LEO. These variables may account for some of the unexplained variation of earnings in the models.

4. Methodology

Analysis focusses on the medium-term outcomes from absence, In two stages: First, the effect of absence on **employment status** (being in sustained employment, on benefits for a sustained period, or on a low income (<£15k p.a.) at age 28. Second, the **marginal effect** absence has on future earnings for those who are employed.

4.1 Employment Status

As being “not employed” is not the same as being “unemployed” (not employed will capture economically inactive), a range of models are required to reflect this distinction.

Being in receipt of benefits for a sustained period of 6 months, being employed and being classed as low income (<£15k p.a.), are binary outcomes. A binomial logistic regression is the most appropriate model to estimate the odds of these outcomes as a consequence of high absenteeism. These models use the following equation (1) to estimate odds ratios of employment status outcomes for absent pupils, while controlling for influential socio-demographic characteristics.

$$Logit(B, E, U)_i = \left(\frac{P_i}{1-P_i}\right) = \beta_1(a_i) + \beta_2 X_i + \varepsilon_i \quad (1)$$

$\boldsymbol{B}$ is the logistic function of receiving benefits at 28, $\boldsymbol{E}$ being employed and $\boldsymbol{U}$ low income <£15k, $\boldsymbol{a_i}$ is the absence variable, and $\boldsymbol{X_i}$ are the sociodemographic control variables, and ε is the error term.

Each model is run twice, with “*total absence*” (continuous variable) and “*persistent absent*” (binary dummy) and the main explanatory variable.

4.2 Marginal Effect of Absence on Total Earnings.

“Total absence” and “*Earnings”* are continuous variables; therefore, an *ordinary least squares regression* (he standard approach in employed by Dräger et al) was selected to explore the marginal effect an increase in absence has on annual earnings at age 28.

This Model uses equation (2), to regress the natural logarithm of “total earnings” on “*total absence rate*” (using the same sociodemographic control indicators as the employment-status models) with the addition of a school-level fixed-effect deployed by Aucejo & Romano (2016).

$$\ln(earnings) = \alpha + \beta_1(a_i) + \beta_2(a_i^2) + \beta_3 X_i + s_i + \varepsilon_i \quad (2)$$

Where the dependent variable is the log of earnings at 28, $\boldsymbol{a}$ is the absence rate for individual i (the focal independent variable). **β1** captures the percentage change in earnings as a result in a 1pp increase in absence. $\boldsymbol{X_i}$ is a set of sociodemographic control variables which can influence earnings, and $\boldsymbol{s_i}$ is the individual’s school unique reference fixed effect. **α** is the intercept (indicating mean earnings at 0% absence) and **ε** is the error term.

5. Mediation Analysis

Human-capital models typically include a schooling variable (the more education, the greater the pecuniary returns), LEO contains *Key Stage 4 points* and *highest level of education* variables. However, it is likely that they are mediators between absence and labour market outcomes:

Absence > GCSEs > Highest level of Education > Earnings

Baron and Kenny (1986) proposed a three-step framework to test if variables are on the causal pathway between the independent and dependent variables, undertaken here:

Step 1 "Total absence" effect on "Total earnings".

Intercept	27228.051 *** (54.219)
Total Absence	-343.981 *** (4.488)
F statistic	5875
Table 3 (standard errors in brackets)	

"*Total absence*" has a significant (5% level) effect on earnings (Table 3). The coefficient suggests a £343.98 reduction in earnings for every one unit increase in absence.

Step 2 "Total absence" effect on "Key Stage 4 points"

Intercept	53.29 *** (0.034047)
Total Absence	-1.03 *** (0.002531)
F Statistic	165,500
Table 4 (standard errors in brackets)	

"*Total absence*" has a significant effect on "Key Stag 4 points" (Table 4). The coefficient suggests a 1.03 reduction in key stage 4 points for every one unit increase in absence.

Step 3a Inclusion of Education Independent variable (KS4 Points).

Intercept	13257.68*** (126.319)
Total Absence	-73.28*** (4.944)
KS4 Points	256.58 *** (2.103)
F Statistic	10480
Table 5 (standard errors in brackets)	

The magnitude of the *"Total absence"* coefficient is reduced drastically (from £343.98 to £72.28 for every one unit increase in absence), however the variable is still significant (at 5% level) (Table 5). Therefore, the inclusion of *"KS4 points" partially* mediates between *"Total absence"* and "*Earnings"*.

Step 3b Inclusion of education independent variable (Attending University or higher).

Intercept	23169.958 *** (66.527)
Total Absence	-228.155 *** (4.578)
University or higher	8509.690 *** (82.596)
F Statistic	8312
Table 6(standard errors in brackets)	

Again, the magnitude of the *"Total absence"* coefficient is reduced when "university or higher" as highest level of education is introduced, therefore it also *partially* mediates between *"Total absence"* and "*Earnings"* (Table 6). The P-Value (<0.00) in a *Sobel test* suggests that the mediation is statistically significant. As a result, both "*KS4 points*" and "*highest level of education*" could be considered mediating variables are therefore **unsuitable as controls**.

Additionally, "*Total absence rate"* and "*Key stage four points"* are moderately negatively correlated (r=-0.5). Highest level of Education and Key Stage 4 points are significantly positively correlated (r=0.7). Which suggests attendance in school has a moderate effect on GCSE results, which has a significant effect on the highest level of education reached. As such, education variables would violate the *Gauss-Markov* assumption on no *multi-collinearity* and could lead to biased estimates of the causal effect of absence on earnings.

6. Empirical Results of Labour Market Outcomes

The following models test the "faucet" and "human capital" theories that poor attenders don't acquire sufficient human-capital to succeed in the labour market (Heckman 2006) as a result of the "faucet being switched off" (Entwisle et al 2001). The first section of these results looks at the relationship between absence and three employment status outcomes. The second section looks at the effect absence has on earnings.

6.1 Employment Status

6.1.1 Impact on being Employed at 28.

	Estimate	Odds Ratio	Std. Error	z value	P Value	CI L	CI H
(Intercept)	1.94	6.97	0.01	283.27	0.00		
Total Absence Rate	-0.04 ***	0.96	0.00	-101.63	0.00	-	-
(Intercept)	1.83	6.20	0.01	279.06	0.00		
Persistent Absence	-0.68 ***	0.50	0.01	-88.35	0.00	0.51	0.49
Table 7							

Table 7 shows the results of estimating equation (1) with being "*Employed at 28*" as the dependent variable. The coefficient of **0.96** points to *"Total absence rate"* (continuous variable) having a significant (at 5% level) reduced effect on the odds of being "*Employed*" (An odds ratio <1 suggests reduced likelihood). For each 1pp increase in absence, the odds of being "*Employed*" are reduced by 4%. For the PA cohort (binary dummy) the odds of being "*Employed*" are reduced by 50% (while controlling for socio-demographic characteristics).
This outcome is consistent with Dräger et al's (2022) conclusion that poor attenders are at greater risk of being "*not employed*". However, this result also captures those who are economically inactive.

6.1.2 Impact on Claiming Benefits For A Sustained Period

	Estimate	Odds Ratio	Std. Error	z value	P Value	CI L	CI H
(Intercept)	-0.40		0.01	-26.47	0.00		
Total Absence Rate	0.02 ***	1.02	0.00	31.97	0.00	-	-
(Intercept)	-0.36		0.01	-24.58	0.00		
Persistent Absence	0.43 ***	1.54	0.01	30.45	0.00	1.51	1.56
Table 8							

Table 8 summarises the results of estimating equation (1) with being "*Claiming Benefits For A Sustained Period At 28*" as the dependent variable. The coefficient of **1.02** points to *"Total absence rate"* (continuous variable) having a significant (at 5% level) increased effect on the odds of being in receipt of benefits.

After controlling for socio-demographic factors known to influence employment outcomes, the odds of being in receipt of benefits increase by 2% for every 1pp increase in "*total absence*" (significant at the 5% level). The odds for the whole PA cohort are increased by 54%.

This result offers a more in-depth insight than Dräger et al (2022) as it excludes the economically inactive. Therefore, this result confirms that those who have greater school absence have higher odds of being “not employed- but seeking work”.

6.1.3 Low income (earning under £15k p.a.)

	Estimate	Odds Ratio	Std. Error	z value	P Value	CL L	CI H
(Intercept)	-2.59		0.01	-291.06	0.00		
Total Absence Rate	0.03 ***	1.03	0.00	64.47	0.00		
(Intercept)	-2.50		0.01	-296.14	0.00		
Persistent Absence	0.54 ***	1.72	0.01	60.29	0.00	1.68	1.75

Table 9

Table 9 summarises the results of estimating equation (1) with “*Low Income* (<£15k p.a.)” as the dependent variable. The coefficient of **1.03** points to *“Total absence rate”* (continuous variable) having a significant (at 5% level) increased effect on the odds of being on a low income at 28.

The odds of earning < £15k in a year increase by 3% for every 1pp increase in “*total absence*” (significant at the 5% level). The odds for the whole PA cohort are increased by 1.72x. Again, this result offers greater insight than existing literature into the negative outcomes for those who are absent.

Additionally, females are more likely to be in this group than males. The raw odds (before controls) of earning less than £15k in a year are 3.2x greater for females than males, which may explain the disparity between their average earnings (*the gender pay gap*). Despite this, the results of this regression point to “low income (<£15k p.a.)” being affected by absence, therefore it cannot be included as a control in any *earnings* regressions due to endogeneity.

These negative outcomes align with human capital theory: Poor attenders are more likely to be either part-time, have a zero-hours contract, or not have held a full-time job for the whole tax year (figure 5).

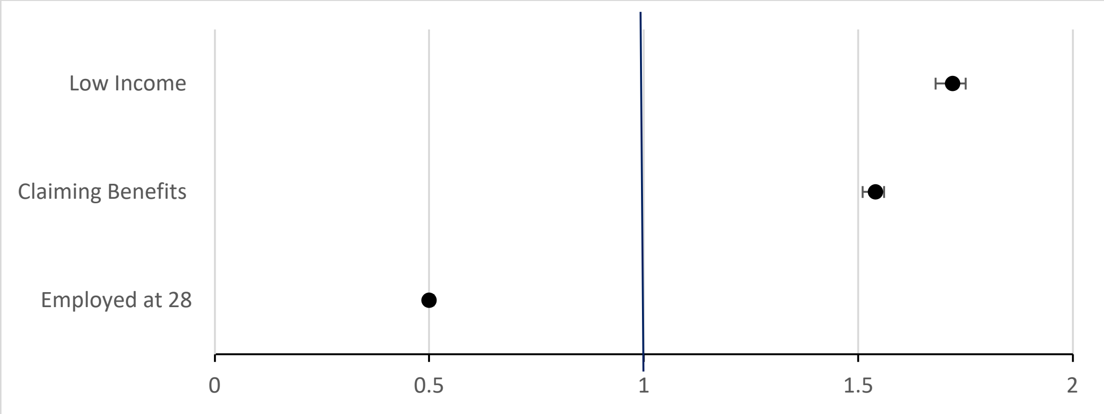

Figure 5 Forest plot demonstrating the odds ratio for Persistent Absent pupils and Employment Status. Source LEO

Robustness Checks
The following tests were conducted to ensure the model estimators were robust.

	ROC Area under the curve	**Wald test**	**log-likelihood**	**Correctly Predicted**
Being Employed	0.63	0.00	0.00	67.6%
Claiming benefits	0.76	0.00	0.00	89.5%
Low Income (<£15k)	0.65	0.00	0.00	86.3%
Table 10				

The **standard errors** in all models are small, indicating a good fit. The **ROC Area under the curve** was greater than 0.6 for "being employed" and "low income", indicating there is some ability to distinguish between true and false results (table 10). However, the "Claiming benefits" ROC value was >0.7 indicating the model is good. A **Wald test** found the independent variables significantly different from 0 for all models. The **log-likelihood** test found the inclusion of absence variables improved all models. Finally, the models correctly predicted a high percentage of the results, with being employed the lowest at 67.6%. As a result, some caution is advised when interpreting the "being employed" model.

6.2 Impact of Absence on Earnings

(Polynomial Regression using Ordinary Least Square method with Fixed Effects)

Table 11 summarises the results of estimating the Polynomial Regression using the OLS method to estimate the marginal effect a*bsence* on *earnings.* The model uses the same time-invariant control variables used previously (to ensure the Gauss-Markov assumption of strict exogeneity is upheld).

	Estimate	Std. Error	t value	P-Value
(Intercept)	10.35 (£31,264.45)	0.090419	114.4694	0.00
Total Absence	-0.03	0.000331	-88.3119	0.00
Total Absence Squared	0.0003	6.43E-06	41.03532	0.00
F-statistic: 19.84 on 3191 and 463776 DF, p-value: < 2.2e-16				
Table 11				

The dependent variable is the "***Log Of Total Earnings"*** (employed and self-employed). The independent variable of interest is the "***Total Absence Rate***".

The sign of the "*total absence*" coefficient in implies the effect of absence on earnings is negative. The estimators point to the marginal effect of a 1pp increase in KS4 absence is a 3% decrease in earnings at 28 (*ceteris paribus*). This result suggests that poor attendance reduces human-capital acquisition, which is reflected in future earnings.

This outcome aligns with the Cattan et al (2022) study but differs from Dräger et al (2022) who were unable to find a statistically significant relationship.

Robustness checks

The **F-statistic** of 19.84 with 3191 degrees of freedom suggests the model has a significant overall fit and the null hypothesis (the variables were not statistically different from zero), can be rejected.

The **t-value** of 114.46, for "*total absence*", is much larger than the critical value of 1.96 (at 0.05 level) and therefore it is statistically significant, and the null hypothesis, that the "total absence" coefficient is equal to zero, can be rejected. The same applies to the "*total absence squared*" variable t-value of 41.

A ***Ramsey Reset test*** initially revealed a mis-specified functional form in the model. This led to using the **natural log of earnings** as dependent variable, to correct for the right skewness of earnings. It was not possible to use the natural log of "***total absence***" as many students had 0% attendance. However, "***total absence squared***" was included in the model (to establish if the relationship was non-linear) and was significant at the 5% level– suggesting non-linearity, however the effect was very small (0.03%) and the curvature of the relationship will be slight. This led to the choice of polynomial regression as the preferred method of estimation.

A ***Breaush-Pagan test*** revealed **heteroscedasticity** was present, which violates the assumption of **homoscedasticity**, (variance in the residuals was not constant as the absence rate increased). As mitigation, the model uses **HAC robust standard errors** (p-value is still <0.05, suggesting the estimator is robust). However, caution is still advised when interpreting the results as there is still potential for bias.

The $\mathbf{r^2}$ was 0.12, suggesting it can account for 12% of the variation in wages. Including endogenous variables may improve the explanatory power, but the estimators would be bias.

6.3 Discussion

6.2.1 Monetising Impact From the OLS Model.

The 3% marginal reduction affects the whole cohort, however "*total earnings*" would depend on the socio-demographic group an individual belongs to.
The mean salary at 0% absence (the intercept) is £31,264. A 3% marginal reduction would roughly equate to a **£938** reduction in yearly earnings per 1pp of KS4 absence.

6.2.2 English As An Additional Language (EAL) Variable.

It was hypothesised that those for whom English wasn't their native language would have reduced outcomes in the medium term (job market). However, the variable was not significant at any level in the model. Additionally, those with an EAL marker (£22,145) earned roughly the same as their English-speaking peers (£22,042) on average.

6.2.3 Endogeneity In Control Variables

Despite having access to a rich dataset there are observable variables, in addition to the education-level variables explored in the mediation analysis, which aren't included in the models to avoid endogeneity problems.

Beyer and Knight (1989) argue that "*occupation*" is equally as important as individual characteristics in wage determination. Job sector would be dependent on their academic achievements and is therefore not exogenous to the model. Further modelling could use **future occupation** as dependent variable and test its association with absence.

There is also a regional disparity in average earnings. House of Commons library Statistics from April 2022 demonstrates median weekly pay was the highest for those living in London (£765) and lowest for those in the North-East (£580) (House of Commons Library, 2022). When "Region at age 28" (the age at which they were earning/in receipt of benefits) was included in the model the intercept and r^2 change to £32,676.60 and 0.13 respectively. The "*total absence*" coefficient was unchanged by the inclusion of region (3%), which adds weight to the argument that "total absence" is a strong predictor of future earnings (as it is unaffected by the addition of a new variable). This also highlights that the impact of absence will be the same relative change (3%) for all groups, however the absolute change in mean earnings will differ depending on the region the individual lives in.

However, a ***Wu-Hausman*** test indicated that the regression was mis-specified when *"Region at age 28"* was included as an instrumental variable. A possible reason for this is that poor attendance may be a barrier to labour factor mobility, making relocating to a region with greater earning opportunities more difficult (the region an individual grew up in will be controlled for by the school fixed effect). The (raw) odds of those living in *London* having at least university level of education are 2.5x greater than those who obtain level 5 or below. As discussed in the mediation analysis – the odds of those who are persistently absent achieving at least a university degree or above are reduced by ~70%. It could therefore reasonably be assumed that getting a degree enables greater labour mobility making "*Region at 28*" an endogenous variable, as such it was removed from the final model.

7.Conclusion

This papers finds poor attendance in KS4 has a statistically significant negative relationship with labour market outcomes. Poor attenders are more likely to be on **benefits**, less likely to be **employed**, and more likely to be on a **low income** (<£15k p.a.) at age 28. Furthermore, increasing absence has a significant negative relationship with earnings at 28. These results are highly meaningful, as it is the first time a statistically significant, monetised figure can be placed on absence in English schools.

Despite these results showing association not causation, it could be reasonably argued the findings are robust. The sample is large (n=566k), data is drawn from a reliable source, detailed socio-demographic characteristics are controlled for, a range of econometric models were used, and corresponding statistical testing took place. In addition, the results of the modelling are consistent with human-capital theory and previous academic research into the impact of absence. Therefore, the inference that absence has a negative effect on labour market outcomes can be made with a degree of confidence.

The findings have similar outcomes to both Dräger et al (2022) and Cattan et al (2022) however the results themselves differed. Like Dräger et al (2022) modelling suggested that the likelihood of being "non-employed" in the medium-term was greater for poor attenders. This study goes further by establishing statistically significant associations with being in receipt of benefits for a sustained period and low earnings too. Unlike the Dräger study, a statistically significant negative-relationship between absence and earnings was also established.

Cattan et al (2022) found 10 days (which would equate to ~5% absence within an English school year) absence at age 10 resulted in a 1-2% reduction in lifetime earnings. This study focussed on reductions in earnings for a single year (age 28) therefore the results aren't directly comparable. Further work could be carried out to extrapolate the results from the model into a lifetime earnings figure.

It should be noted that this study uses absence at a different age to both papers (16 compared to 10) and covers a longer period than the Dräger et al (2020) study (five and a half terms compared to one). The data may also be more accurate than the Cattan study, given it is more recent and comes directly from government records.

There is still potential for an unobserved factor (some of which are listed in the limitations section) which may either explain the variation in earnings or negative labour market outcomes, which hasn't been controlled for in the models. As a result, a degree of caution is advised when interpreting these results. It is unclear if poor attendance is a key driver of diminished short-and-medium-term outcomes, or if absence, academic achievement, and labour market outcomes are all the symptoms of an unobserved variable such as poor work ethic or lack of aspiration.

Caution is also advised when applying these results to the current unprecedented levels of absence as the results within are based on attendance and earnings in a typical year. The Covid-19 pandemic has had a profound effect on school attendance and the reasons for absence may differ (e.g., mental health issues), therefore the outcomes for future generations may vary. As a result, analysis on cohorts who were affected by the 2020 school closures should be carried out – to see if their academic outcomes differ from the cohort in this study.

7.1 Policy Recommendations.

The results of this research are alarming, given the current high levels of absence. The following policies are recommended:

- Place a greater emphasis on good attendance by setting a target of bringing **absence levels** pupils back to within the pre-pandemic range. Negative outcomes could be avoided by incentivising attendance with financial rewards for improved attendance (such as vouchers as deployed in Deyton, Ohio (Fortin, 2022)). Alternatively use the *"£938 reduction in yearly earnings"* figure prominently to disincentive absence.

- Interventions targeting the small group of persistently absent pupils could have a sizable effect. Despite making up only 30% of the cohort, PA pupils accounted for 65% of all KS4 absence sessions.

- Initiate policies that remove the barriers to attendance by investing more in **mental health provision** and introduce tailored interventions for those who have become persistently absent since the pandemic.

- Target pupils from disadvantaged backgrounds (FSM and SEN) as a priority. These individuals have higher absence rates which could explain their diminished outcomes in the labour market. Those in both socio-demographic groups **(Free School Meal & Special Educational Needs)** have higher than average absence (17.0% compared to 9.2%), lower average KS4 Points (19.0 compared to 43.8) and lower average yearly earnings (£16467.63 compared to 24,334.81).

References

Baron, R. M., & Kenny, D. A. (1986). **The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations**. Journal of Personality and Social Psychology, 51(6), 1173–1182. Available at (https://doi.org/10.1037/0022-3514.51.6.1173) (accessed 04/05/2023)

Becker (1964, 1993, 3rd ed.). "**Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education**"

Beynon K & Thomson D, (2022) **To what extent is the KS2 disadvantage gap explained by pupil absence?,** available at (https://ffteducationdatalab.org.uk/2022/09/to-what-extent-is-the-ks2-disadvantage-gap-explained-by-pupil-absence/) (accessed 10/01/2023)

Beynon K & Thomson D, (2021) **How many primary school pupils have been persistently absent so far this year?** Available at (https://ffteducationdatalab.org.uk/2021/11/how-many-primary-school-pupils-have-been-persistently-absent-so-far-this-year/) (accessed 10/01/2023)

Cattan. S, Kamhofer. D, Karlsoson. M, Nilsson. T. (2022) **The Long-Term Effects Of Student Absence: Evidence From Sweden**. The Economic Journal, 133 (February), 888–903 https://doi.org/10.1093/ej/ueac078

Card, David. (1996) "**The Effect of Unions on the Structure of Wages: A Longitudinal Analysis**." Econometrica, vol. 64, no. 4, 1996, pp. 957–79. Available at (https://doi.org/10.2307/2171852). (Accessed 31/03/2023).

Chang, H, Romero, M (2008) **Present, Engaged, and Accounted For The Critical Importance of Addressing Chronic Absence in the Early Grades** Available at (http://www.nccp.org/wp-content/uploads/2008/09/text_837.pdf) (accessed 12/12/2022)

Department for Education (2016) **The link between absence and attainment at KS2 and KS4**. Available at (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/509679/The-link-between-absence-and-attainment-at-KS2-and-KS4-2013-to-2014-academic-year.pdf) (accessed 13/12/2022)

De Beyer. J, Knight. J. B, (1989) **"The Role Of Occupation In The Determination Of Wages"**, *Oxford Economic Papers*, Volume 41, Issue 1, 1989, Pages 595–618, https://doi.org/10.1093/oxfordjournals.oep.a041916

Department for Education (2018) **Outcomes for pupils eligible for free school meals and identified with special educational needs**. Available at: (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730977/FSM_and_SEND_outcomes-statistics.pdf) (accessed 07/03/2023)

Drager. J, Klein. M, Sosu. E, (2022) The long-term consequences of early school absences for educational attainment and labour market outcomes, available at (https://pure.strath.ac.uk/ws/portalfiles/portal/155044825/Drager_etal_EdArXiv_2023_The_long_term_consequences_of_early_school_absences_for_educational_attainment.pdf) accessed (04/05/2023)

Education Endowment Foundation. (2018). **Sutton Trust-EEF Teaching and Learning Toolkit & EEF Early Years Toolkit: Technical Appendix and Process Manual**. Available at (https://educationendowmentfoundation.org.uk/public/files/Toolkit/Toolkit_Manual_2018.pdf) (accessed 10/01/2023)

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2001). **Keeping the Faucet Flowing: Summer Learning and the Home Environment**. American Educator, 25, 10–15.

Explore Education Statistics (2023) Available at (https://explore-education-statistics.service.gov.uk/data-tables) (accessed 08/03/2023)

Fortin, J. (2022**). More Pandemic Fallout: The Chronically Absent Student**. New York Times. Available at (More Pandemic Fallout: The Chronically Absent Student - The New York Times (nytimes.com)) accessed (03/05/2023)

Folguera, C, Lamolla, L, Marin, X (2022) **Working-class women who work part-time: choice or necessity?** Available at (https://dobetter.esade.edu/en/women-work-part-time#:~:text=The%20reasons%20given%20for%20choosing,adults%20is%20the%20main%20factor) (accessed 07/04/2023)

Glass, G. V., B. McGaw, and M. L. Smith. (1981). **Meta-Analysis in Social Research**. Beverly Hills, CA: Sage Publications.

Greenwood, C., Jenkins, A. and Vignoles, A. (2007) **The Returns to Qualifications in**

England: Updating the Evidence Base on Level 2 and Level 3 Vocational Qualifications. CEE Discussion Papers 0089. Centre for the Economics of Education, LSE.

Hodge, L., A. Little, and M. Weldon. 2021a. **"GCSE Attainment and Lifetime Earnings**." Research Report. Department for Education. Available at (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993202/GCSE_Attainment_and_Lifetime_Earnings_PDF3A.pdf) accessed (20/01/2023)

House of Commons Library. (2002) "**Average Earnings By Age And Region**". Available at (https://commonslibrary.parliament.uk/research-briefings/cbp-8456/#:~:text=Regional%20differences,North%20East%20at%20April%202022.&text=The%20East%20and%20South%20East,%E2%80%9Cin%20real%20terms%E2%80%9D) accessed (30/03/2023).

Heckman, J; Lochner, L; Todd, P (2003) "**Fifty Years Of Mincer Earnings Regressions**", Available at (https://www.nber.org/system/files/working_papers/w9732/w9732.pdf) (accessed 31/03/2023).

Lázaro, S.; Urosa, B.; Mota, R.; Rubio, E. (2020) **Primary Education Truancy and School Performance in Social Exclusion Settings: The Case of Students in Cañada Real Galiana**. *Sustainability 12*, 8464. Available at (https://doi.org/10.3390/su12208464) (accessed 12/12/2022)

Mcintosh, S. (2006) **Further Analysis of the Returns to Academic and Vocational Qualifications**, Oxford Bulletin of Economics and Statistics, 68(2), pp. 225–251

Sims, S. (2020). **School Absences and Pupil Achievement**. CEPEO Briefing Note No. 1. London, UK: Centre for Education Policy and Equalising Opportunities, UCL Institute of Education.

Tversky. A, Kahneman. D, (1991) **Loss Aversion in Riskless Choice: A Reference-Dependent Model.** The Quarterly Journal of Economics, Volume 106, Issue 4, November 1991, Pages 1039–1061, available at (https://doi.org/10.2307/2937956)