Department for Education

The Impact of School Absence on Lifetime Earnings

Research report

March 2025

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
KS2, KS3 & KS4	Key Stage 2 – ages 7-11 (primary school) Key Stage 3 – ages 11-14 (secondary school: years 7-9) Key Stage 4 – ages 15-16 (secondary school: years 10-11)
PA	Persistently Absent – missing more than 10% of potential sessions
SA	Severely Absent – missing more than 50% of potential sessions
SD	Standard deviation – the average distance from the mean.

Acknowledgements

This analysis was undertaken by economists at the Department for Education (DfE). We would like to acknowledge the extensive assistance from DfE colleagues, notably Tobin Lochrie, William Gobey, Hannah Tarrant, Alex Davenport, Chris Cook, and Andrew Luty.

Additionally we would like to thank Arpita Bhattacharjee, Megan Arnot, Bethany King, Harry Strange, Grant Davies, Reece Fraser, Caroline Halls and DfE Science Advisory Council member Claire Crawford for reviewing and quality assuring the analysis and report.

Executive summary

This report estimates the monetary impact of one day of school absence within state-funded English secondary schools. First, we model the association between absence in Years 7–11 and Key Stage 4 attainment. Then, we apply these results to previous departmental research on the lifetime earnings returns to education. Combining these results enables us to estimate the effect of increased absence from school on reduced future earnings for individuals.

The findings are consistent with previous departmental research in this area, which found pupils who had between 95-100% attendance in Year 11 were 1.9 times more likely to achieve a Grade 5 in English and Maths GCSE compared to pupils who only attended 90-95% of the time (DfE, 2025).

The motivation behind our analysis is to obtain a monetised figure which could be included in cost benefit analysis calculations for policies aimed at reducing school absenteeism or where school attendance might be impacted. Additionally, the estimates within this paper can also underpin discussions on the importance of attending school, highlighting how this can affect skills and human capital in the economy.

We caution against overly strong causal interpretations of these findings. Our results show the association between absence and KS4 attainment, by controlling for as many factors that could affect these outcomes as we can. Nonetheless, our results do show clearly that higher absence is associated with worse attainment and therefore worse earnings outcomes in later life. The findings in this report are also of a similar magnitude to those in existing studies that do suggest a causal relationship. If these findings are used in cost benefit analysis, they should be accompanied by the caveats contained within this report.

Key Findings

We find a negative relationship between school absence and attainment; a pupil who misses more days of school is likely to achieve worse grades in their GCSEs. This finding is consistent with previously published literature.

On average, one day of absence between Years 7 to 11 is associated with:

- A reduction of **0.68%** of a standard deviation in Attainment 8[1] scores at KS4 for a pupil with the average level of prior absence[2]. This approximately equates to a 1 grade decrease per 13 days of absence[3].

[1] One standard deviation in the sample is 11.2 GCSE grades.
[2] See methodology section
[3] 0.68% (effect size) x 11.2 (1 SD) = 0.076 grades , 1/0.076 = 13.1

We use these results to estimate the *indirect* effect on earnings, by applying the result to the 'GCSE attainment and lifetime earnings' estimates previously published by the Department for Education[4].

Using this method we estimate:

- One day of additional absence between Years 7 to 11 for a typical student was associated with an approximate £750[5] (2024 prices) loss in future earnings, discounted to present value terms[6].
- One day of absence for persistently absent pupils, who miss more than 10% of their possible sessions, was associated with a £650[7] future earnings loss (2024 prices, present value discounted terms).

We also use a different dataset, which links earnings and academic records, to test if there is a *direct* association between KS4 absence and a range of future labour market outcomes.

- We find a one day increase in absence in Years 10 and 11 is associated with a **0.8%** decrease in total yearly pay-as-you-earn (PAYE) earnings and declared self-employed earnings at age 28[8].
- We find the likelihood of being in receipt of benefits increases by **2.7 times** for pupils who are classified as persistently absent (>10% absence). This rises to **4.2 times** for those who are classified as severely absent (>50% absence).
- The likelihood of being in sustained employment for 12 months decreases by approximately **60%** for pupils who are classified as persistently absent and approximately **75%** for those who are classified as severely absent.

[4] Lifetime earnings report [GCSE attainment and lifetime earnings (publishing.service.gov.uk)]
[5] Effectively the 46th day of absence between years 7-11
[6] Throughout the report, we refer to 'lifetime earnings' as shorthand. This is 'lifetime earnings in present values', discounting lifetime earnings values to the present day using standard HM Treasury discount rates of 3.5% over years 1 to 30, and 3% thereafter (HM Treasury, 2020).
[7] Effectively the 90th day of absence between years 7-11
[8] These figures to be used in narrative/context setting only – not to be used in cost benefit analysis.

Introduction

The percentage of pupils who are regularly absent from school has risen since the COVID-19 pandemic-related school closures. Prior to the pandemic, the overall absence rate, missing school for either authorised or unauthorised reasons, was **4.7%** in the 2018/19 academic year. Once schools began to reopen, following the easing of restrictions, absence rates rose sharply to **7.6%** in 2021/22 academic year, and have remained high, only falling by 0.2pp to **7.4%** in the 2022/23 academic year [9].

Furthermore, the percentage of pupils who are classed as **persistently absent**, missing more than 10% of possible school sessions, rose from **10.9%** in 2018/19 to **22.5%** in 2021/22. This rate only fell to **21.2%** in 2022/23. The percentage of **severely absent** pupils, those who missed greater than 50% of possible sessions, more than doubled from 2018/19 when the rate was **0.8%** to **2.0%** in 2022/23[10]. These figures are concerning, as the consensus amongst the literature points to absence having a negative relationship with both school attainment and labour market outcomes.

We use two approaches to estimate the association between absence and earnings in this report, both are valued for difference reasons. In Section One we present our indirect analysis which assumes school attainment is a mediation mechanism between absence and earnings; missing school results in lower GCSE grades which then affects future employment prospects. Section Two presents our direct approach, which looks at the relationship between absence and yearly earnings for one tax year, when individuals are aged 28. Additionally we establish the odds associated with different absence levels and certain labour market outcomes.

In Section One, we begin by modelling the association between attendance and Attainment 8 scores. Absence data is taken from state-funded secondary school pupils between Years 7 to 11. As absence is not the only issue likely to affect a pupil's attainment, we also account for: prior attainment, eligibility for Free School Meals (FSM), Special Educational Needs (SEN) status, sex and month of birth and the differences between schools. Despite these controls, we cannot say these results are deterministic. We can't fully capture pupils' backgrounds and individual circumstances, and we do not assert that pupils who miss school would automatically achieve higher attainment if their attendance was higher.

Previous research by the department has been able to estimate earnings trajectories based on GCSE results. We exploit this research to estimate lifetime earnings associations with the changes in attainment correlated with absence. The resulting output is a monetised lifetime earnings figure which can be used to facilitate the cost benefit

[9] Explore Education Statistics: https://explore-education-statistics.service.gov.uk/data-tables/permalink/371f1e4b-fdb1-4767-93cb-08dcafcfd28b
[10] Explore Education Statistics: https://explore-education-statistics.service.gov.uk/data-tables/permalink/45d5ce93-f99f-470a-9413-08dcafcfd28b

analysis of policies and legislation aimed at improving attendance in state-funded schools. This approach is important for Department for Education policymakers, as it enables us to assess the value for money of policies which aim to improve both attendance and attainment. Additionally, this analysis can provide context to underline the importance of equipping young people with the skills and knowledge to give them the best chances in the labour market.

The second section of this report uses earnings data to establish if there is a direct association between school absence and income in one tax year, when the individual is aged 28. We use the Longitudinal Educational Outcomes (LEO) dataset for this analysis, which links education, tax and benefits data from Department for Education (DfE), His Majesty's Revenue and Customs (HMRC), and the Department for Work and Pensions (DWP). As this analysis considers earnings during one tax year in isolation, it is intended to complement the main findings from Section One and can be used in any context setting or narratives to underline the importance of good attendance.

Table 1: Uses of Our Two Approaches.

Approach	Provide context in written narratives	For use in Cost Benefit Analysis	Stage of Earnings
Indirect analysis.	✓	✓	Lifetime
Direct analysis	✓	X	One tax year at age 28

Caution is advised when interpreting any findings within this report. We find absence has a negative correlation with GCSE attainment and earnings. There is uncertainty around the causal mechanisms, so the results should be interpreted as associations rather than fully causal. However, the literature review section of this report highlights numerous rigorous studies which have consistently also found a negative relationship between absence and attainment across a range of schools, students, and locations. These studies use a range of statistical methods, some of which point to causal relationships (Aucejo & Romano, 2016; Gaete-Romano, 2018; Liu et al, 2021). However, if the figures from this report are used in a cost benefit analysis the caveats included within this report should accompany them in the written outputs.

Furthermore, context should be considered when quoting these estimates. We interpret our findings as applying to the effect of uncoordinated absence: individual pupil absence when their school is open. The results are therefore not directly applicable to coordinated absences (when the whole school is closed due to industrial action, extreme weather or COVID-19). However, we haven't stripped out coordinated absence due to data limitations. As uncoordinated is much more prevalent we use these results as a proxy for uncoordinated absence. Coordinated absence may be expected to have a lesser effect, as schools take mitigating action to minimise class-wide learning loss.

We provide estimates for pupils who have an average level of prior absence and also for persistently absent pupils, which we define as those who miss more than 10% of possible sessions. We do not provide estimates for the impacts of absence for severely absent pupils, defined as those who miss more than 50% of schooling. This decision reflects that these pupils are outliers and require separate analysis. As such, these results apply to pupils who are not classed as severely absent.

Literature Review

There are a number of academic papers which estimate the impact of absence on attainment, using a range of econometric methods. There are fewer papers on the effect of absence on labour market outcomes. These studies span both primary and secondary age pupils across a range of different countries.

Here we summarise the findings most relevant to this report, beginning with impacts on education, then moving onto the impact on labour market outcomes.

Impact of school absence on attainment

The table below lists the literature on attendance and attainment we reviewed as part of this report. The findings are presented as *effect sizes*, or percentage of a standard deviation (SD). Effect sizes allow us to robustly compare the impact across cohorts, location, and time. We also present the year the study was conducted, the age of the data used (where given), their method to estimate the impact, the country or US state of the study, the age of the pupils who are the subject of their analysis, and the study’s sample size. We then proceed to discuss the results and limitations in more detail below. Some of the papers report the effect size per day of absence, while others have reported by percentage point (ppt) increase. We have highlighted this distinction in the relevant column.

Table 2: Literature on Absence and Attainment.

Paper	Year	Method	Country	Effect size	Age of Pupils	Sample Size (n)
Sims	2020	Meta Analysis	Various	**0.3 - 0.4** % of a SD per day.	- (multiple)	- (multiple)
Klein, Sosu, Dräger, Casoni	2024	Multiple (this figure OLS)	UK	**6.4%** of a SD per **1ppt** increase	UK year 1-11	8,139
Klein, Sosu, Dräger, Casoni	2024	Multiple (this figure OLS)	UK	**1.9%** of a SD Maths and **1.6%** of a SD in English per **1ppt** increase	UK year 1-11	8,139
Esteban M. Aucejo a , Teresa Foy Romano	2016 (Absence data from 2006-2010)	Student and School Fixed Effects	USA North Carolina	**0.55%** of a SD *Maths* **0.29%** of a SD *English* per day	US Grade 3-5 Age 8- 11	1,302,037
Esteban M. Aucejo a , Teresa Foy Romano	2016 (Absence data from 2006-2010)	Flu Cases as Instrumental Variable	USA North Carolina	**1.02%** of a SD Maths **1.82%** of a SD English per day	US Grade 3-5 Age 8- 11	1,302,037
Cattan, S. Kamhofer, D. Karlsson, M. Nillson, T.	2022 (Absence data from 1937–47)	OLS, Individual and Sibling Fixed Effects	Sweden	**0.45- 0.51**% of a SD per day	Swedish Grades 1 and 4 (up to age 11)	17,999
Gaete-Romeo	2018 (Student strike data: 2011/12)	Student strikes as an instrumental variable	Chile	**0.95%** of a SD for Maths per **1 ppt** increase in absence.	Chilean High School	1,033,404
Gershenson, Jacknowitz, and Brannegan	2015	Value Added Model with year, grade, teacher, and school FE	USA North Carolina	**0.2-0.7%** of a SD Maths and **0.2 - 0.4%** of a SD for Language, per day	Primary age US Grade 3-5 Age 8- 11	587,919
Gershenson, McBean, and Tran	2018 (Absence data from 1980s)	Value Added Model with year, grade, teacher, and school FE	USA Tennessee	**0.7%** of a SD per day Maths	Primary age US Grade 1-3 Age 8- 11	6,500

Paper	Year	Method	Country	Effect size	Age of Pupils	Sample Size (n)
Goodman	2014 (Absence data from 2003-2010)	Fixed Effects	USA Massachusetts	***FE*- 0.8%-2%** of a SD Maths, **0.8%-1.6%** of a SD English Language per day	Primary age US Grades 3-8 and 10	19,846
Goodman	2014 (Absence data from 2003-2010)	Snow Days as Instrumental Variable	USA Massachusetts	***Uncoordinated** absence* ***IV*** - 5.3% of a SD (Maths), 1.2% of a SD (English Language) per day [insignificant]	Primary age US Grades 3-8 and 10	19,846
Goodman	2014 (Absence data from 2003-2010)	Snow Days as Instrumental Variable	USA Massachusetts	***Coordinated** absence* ***IV*** - 0-0.5% of a SD (Maths) [insignificant], 0.3% of a SD (English Language) per day [insignificant]	Primary age US Grades 3-8 and 10	19,846
Gottfried	2011 (Absence data from 1994-2000)	Family Fixed Effects	USA Philadelphia	13% of a SD per day (Maths), 9.1% of a SD per day (Reading)	Primary 6-13	*Full Sample* 33,400 *Sibling Sample* 6,872
Liu, L. Lee, M. Gershenson, S.	2021 (Absence data from 2002-2003)	Fixed Effects value added models	USA California	0.3 - 0.4% of a SD per day	Secondary	70,000

Discussion of Findings and Effect Sizes

Meta analysis by Sims (2020) for the Centre for Education Policy & Equalising Opportunities at UCL, summarises much of the literature listed above. They estimate each day of individual pupil absence results in a reduction in attainment of approximately **0.3-0.4%** of a standard deviation (SD).

However, there is a large degree of variation between each study's effect size. Gershenson et al (2015) estimate the effect on maths is between **0.2%** and **0.7%** of a SD per day whereas Goodman (2014) estimates the range is between **0.8%** and **2.0%** of a SD. Both studies are based on primary age pupils within the US education system. Furthermore, Gottfried (2011) finds one day of absence is associated with an effect size of **13%** of a SD for Maths, however this effect size appears to be an outlier when considered alongside the other findings. There is a similar level of variation in the results for English or Reading, albeit with smaller effect sizes.

A range of studies (Gaete-Romeo (2018), Gershenson, Jacknowitz, and Brannegan (2017), Goodman (2014), Gottfried (2011), Liu, Lee, and Gershenson (2021)) all found absence has a larger impact on Maths attainment than on English Language or Reading. This suggests pupil performance in different subjects is determined by specific skills, some of which might be harder to learn outside of school.

Goodman (2014) indicates coordinated absence, when a school is closed to all pupils, results in smaller effects than uncoordinated absence. This may be due to schools being conscious of missed learning and may proactively seek to remediate, whereas uncoordinated absence may go untreated. The results of the study were not statistically significant though. Sims (2020) concludes there is a relative lack of studies on coordinated absence, and it is unclear whether coordinated absences are more or less damaging than uncoordinated absence.

Klein et al (2024) conclude school absence harms both short-term school attainment and longer-term labour market outcomes for British individuals. They found a 1% increase in absence resulted in 6.4% of a standard deviation reduction in attainment. Assuming a 190-day academic year, this converts into a reduction of approximately 3.4% of a SD per day of absence.

Methodologies

There are a range of methodologies employed by each paper. Gershenson et al (2015 and 2018) and Liu et al (2021) are able to use *value added models* as their datasets contain attainment results in short intervals (at the end of each year). This allows for a recent baseline of prior attainment to be included and eliminates some of the unobserved factors that could influence attainment, through the first difference principle. However, the low frequency of standardised assessments in English schools makes this model unviable for our analysis, given formal testing is only conducted at the end of Key Stages 2 and 4.

To account for heterogeneity in their models, Aucejo and Romano (2016), Cattan et al. (2022), Gershenson et al (2015) Goodman (2014) Gottfried (2011) and Liu et al (2021) use individual, family, or school fixed effects in their models. Our data set allows us to include school-level fixed-effects in our models.

Some papers use instrumental variables to imply a causal relationship between absence and attainment. Aucejo and Romano (2016) use flu cases, Gaete-Romeo (2018) uses strike data, and Goodman (2014) uses snow days. Only the Goodman study was unable to establish a statistically significant relationship. To note, these variables are not recorded in our dataset, and we are unable to use them in our analysis.

Linearity assumption

Many of the studies, (Gershenson, Jacknowitz, and Brannegan (2017); Liu, Lee, and Gershenson (2021); Cattan et al (2017)) assume the relationship between days of absence and attainment is linear. This implies that a pupil missing 10 days of schooling will have 10 times the effect of missing 1 day.

In our analysis, we explore the possibility of non-linearity and examine whether one day of absence has more impact for pupils with low prior absence than those who have already missed large amounts of schooling. After testing both linear and non-linear models, we conclude that the effect of school absence appears to be non-linear. This relationship can also be visually seen in Figure 1.

Many of the studies only use one academic year of absence, whereas our analysis covers a five-year window, between Years 7 and 11. This larger dataset of absence allows us to observe non-linearities in the data which may not be apparent in single-year analysis. We hope this analysis contributes to the evidence base on non-linear effects of schooling and encourages further analysis of this topic.

Impact on Labour Market Outcomes

There are fewer studies that link marginal changes in attainment to whole life-course labour market outcomes. The two notable papers that do investigate this link, Cattan et al (2022) and Klein et al (2024), both find school non-attendance to be associated with negative future labour market outcomes.

Cattan et al (2022) observed a link between absence and the future earnings of Swedish individuals born between 1930 and 1935. Their analysis found 10 days absence in an academic year could result in lower lifetime earnings of between 1%-2%.

Additionally, Klein et al (2024) discovered a statistically significant link between school absences and non-employment. Specifically, individuals who missed five days of school at age 10 were 0.6 percentage points more likely to be non-employed compared to those who had never been absent. However, they found no statistically significant relationship between absence and sustained unemployment or future earnings once they controlled for various other factors known to influence labour outcomes.

Data Sources

Cattan et al (2022) use pension and tax data from 1962 and 1972 population census records to calculate lifetime earnings. This was then joined to the individual's absence data, taken from 1940s school records.

Klein et al (2024)'s attainment models use data from the Millennium Cohort Study[11] a panel dataset which follows around 19,000 individuals, linked to the 2015/16 cohort in the National Pupil Database (NPD), the same database our study uses. Like ourselves, they have access to a long period of absence data to use in their modelling.

For their analysis of labour market outcomes, they use the 1970 British Cohort Study (BCS70)[12], a panel dataset which follows around 17,000 people born in a single week of 1970 in the UK. The BCS70 includes survey information from teachers on pupil absences at age ten and labour outcomes between the ages of 32 and 42. This has the limitation of only collecting absence for one school year, which would be Year 6 in current academic year labelling.

Methodologies

Cattan (2022) uses within-family (between siblings) fixed effects in their models to control for family level heterogeneity. They caveat that there are still many unobservable variables that may impact school absence (uninspiring teachers, poor health) which may lead to spurious correlations between absence and labour market outcomes.

Klein et al. (2024) use regression-based approaches, such as OLS regression, linear probability models, and two-way fixed effects regressions, to examine the short- and long-term effects of pupil absences on school achievement, education, and labour market outcomes. This method allows the researchers to control for individual heterogeneity and observe the effects over time. We make use of a panel dataset in the second section of this analysis to observe the direct association between absence and labour market outcomes at the earlier age of 28.

[11] MCS: Millennium Cohort Study - CLOSER
[12] BCS: CLS | 1970 British Cohort Study

Limitations

A significant limitation of all the papers in the literature review, for UK policy making purposes, is that all but one are conducted outside the UK. Most are based on students in the United States, with one study coming from Chile (Gaete-Romeo, 2018), one from Sweden (Cattan et al, 2022) and one from the UK (Klein et al 2024). There is cultural and social heterogeneity between these countries and the UK, which makes selecting an effect size from the non-domestic studies problematic for UK policymakers.

Furthermore, some of these findings are over ten years old (Gottfried, (2011), Goodman (2014)). Cattan's (2022) study, based in Sweden, uses absence data from 1937-1947. The most recent study is Klein et al (2024) which uses a range of data sources but with smaller samples than we analyse.

Finally, many of the papers, including both labour market papers, focus exclusively on primary age pupils. The analysis in this paper uses secondary age pupils only, where the effect of absence may differ from primary school pupils. Published data from the Department for Education shows absence in secondary schools is much higher than primary schools[13]. There are a variety of explanations for this. However, we have reservations around applying the conclusions from studies on primary schools to absence in English secondary schools.

[13] Explore Education Statistics: https://explore-education-statistics.service.gov.uk/data-tables/permalink/e6efe1b4-8871-48b3-db41-08dd42ed7adb

Section One – Indirect Approach

Data Source

This section of the analysis uses data from the National Pupil Database (NPD) [14], an administrative dataset that contains a range of anonymised data about every child in the English state-funded school system. We do not include pupils who attended independent schools in this analysis.

Personal characteristics and absence records are collected via the school census. Key Stage 4 attainment data is collected from the awarding bodies and joined to the census data. It should be noted there is an accepted margin of error for the joining of census and attainment data in the NPD which may result in a small percentage of mismatches.

Cohorts Used in this Analysis

Conducting analysis on pupils from just one academic year could raise questions around the ability to apply the findings to future cohorts. Therefore, we conducted analysis on a dataset comprised of the three cohorts who took their GCSEs in academic years 2016/17, 2017/18 & 2018/19 giving us a sample size of 1.4m pupils[15].

During this period, English schools transitioned from *letter grades (A-G) to number grades (9-1)** for GCSEs in a phased approach:

- **2017:** First subjects (English Language, English Literature, Maths) switched to 9-1 grading.
- **2018:** More subjects, including Sciences, Geography, History, and Modern Foreign Languages, adopted the new system.
- **2019:** Nearly all GCSE subjects used 9-1 grading.

Number grades do not directly match the previous letter grades. For example, a Grade 9 is a higher grade than an old A*. Grade 4 is the minimum "standard pass", while Grade 5

[14] We don't use the Longitudinal Education Outcomes (LEO) dataset in this analysis as it only contains absence records from 2005/06 onwards, meaning even if we restricted attention to absences in Years 10 and 11 only (as we do in Section 2), we can only observe labour market outcomes at age 28 for these individuals. We instead apply the findings from the NPD to previous research by the Department for Education, which estimates the lifetime earnings returns to GCSE attainment.

[15] Annex A demonstrates there is little variation between effect sizes when the same regression is run using the single cohorts separately.

is a "strong pass". The absolute grades may not be directly comparable; however we standardise them, which should negate this problem.

Pupils in our dataset have uninterrupted attendance data between Year 7-11. Cohorts from GCSE year 2019/20 onwards could reasonably be considered non-typical. These pupils have a combination of incomplete absence records and/or teacher-estimated GCSE grades, as a result of the COVID-19 disruptions. We took the decision not to include these pupils in our modelling as their records may affect the results.

The software filters out records which have missing data. For example, there were 577 records in the dataset with no ethnic group recorded. These were removed from the modelling where ethnic group was controlled for. Missing data was the only reason for filtering the dataset and improbable values were not removed (see Annex C).

Pupil Characteristic Variables

Previous research has identified certain pupil characteristics which correlate with attainment. We account, or control, for these characteristics in the modelling. The details of these variables are listed below:

Table 3: Pupil Characteristic Variables

Characteristic Variable	Description
Ethnicity	Minor Ethnic group is included in the model as a categorical variable, with *White British* as the reference group. Pupils without a record for ethnicity are removed from the modelling.
Gender	Gender, as recorded in the school census, is included as a categorical binary-variable[16] separating: • Boys (reference group) • Girls
Free School Meal (FSM) status	FSM eligibility is included in the model as a (binary) categorical variable separating: • Not eligible for free school meals (reference group) • Eligible for free school meals (in year 11)
Special Educational Needs (SEN) status	SEN status is included in the model as a (binary) categorical variable separating: • No identified SEN (reference group) • At least 1 identified SEN (Education Health Care Plan or SEN support).
Month of Birth	Month of birth is included in the model as a categorical variable.
School (Fixed effect)	The school's unique reference number (URN), when the pupil is in Year 11 is included as a fixed effect categorical variable. The school at which the pupil achieved their prior attainment is not included.
Cohort (Fixed effect)	The individual's cohort (GCSE year) is included as a dummy variable to control for heterogeneity between different academic years.

School Fixed Effects

We include pupils' Year 11 school Unique Reference Number (URN) as a school-level fixed effect (FE), similar to the Aujeco (2017) and Goodman (2014) specifications. Running a model with school fixed effects addresses the omitted variable bias which arises from the heterogeneity across schools, such as leadership, culture and peer effects. This also then controls for selection into treatment bias – as school allocation isn't random.

[16] This is consistent with other DFE publications.

It also improved the goodness of fit compared to other specifications we tried. This method ensures the results are generalisable, regardless of the specific school a pupil attended. However, it does not fully eliminate the possibility of omitted variable bias from unobserved individual, class or teacher variation.

Attainment Variables

We use three measures as dependent variables: Attainment 8[17], GCSE Maths score, and GCSE English Score.

- **Attainment 8** is calculated by adding together pupils' highest scores across eight approved subjects taken from the following three categories:

Category 1 – *English* and *Maths*, worth double marks, English only counts double if both Literature and Language are taken. The higher grade of the two is used.

Category 2 – Top three scores from the English Baccalaureate (EBacc) subjects taken, i.e. Sciences, Computer Science, History, Geography and Languages.

Category 3 – Top three scores from remaining EBacc subjects or other approved qualifications i.e. other GCSEs or Level 2 Certificates in some technical subjects.

These are then converted into numerical values.

- **GCSE Maths** and **English** are included using the new grading system, which runs from 9-1, with 9 being the highest and 1 being the lowest grade. The previous system graded students from A*-G (with some unclassified – U).
- **KS2 attainment** is a continuous variable and used as a control for individual heterogeneity in the modelling. We use this to control for prior attainment. This variable is a composite of SATs Reading mark + SATs Maths mark.

To adjust for differences between academic years in grading standards, examination difficulty, and grade variation, we *standardise* attainment scores to create a variable that is consistent across cohorts and time. This involves taking away the mean score and dividing by the standard deviation for each cohort. We therefore report an 'effect size' or percentage of a standard deviation, rather than a reduction in GCSE or Attainment 8 points.

Overall Absence Variable

The main explanatory variable of interest is overall absence days between Years 7 to 11. Absence data is collected from the termly school census and converted into days for modelling purposes. We use absence from autumn, spring, and summer terms for Years

[17] We don't use Progress 8 scores as we control for prior attainment.

7-10 and then autumn and spring for Year 11. We omit the final summer term due to different schools starting study leave for pupils at different points during the pupil's GCSE year, which has the potential to skew the modelling. See Annex D for further explanation of how we calculated number of days.

For modelling purposes, we don't separate out authorised or unauthorised absence.

A combination of the following makes disentangling absence by reason difficult:

- *School error*: the school/teacher records the reason for absence incorrectly in error.
- *Disingenuous reporting*: the reason reported is not the genuine reason for absence i.e. a parent reports a holiday during term time as illness.
- *Lack of reporting*: the family of the pupil fails to provide a reason. This will be recorded as unauthorised, but the genuine reason may be illness.

Furthermore, it is unlikely a pupil exclusively has only authorised or unauthorised absences on record, rather a combination of the two. Modelling the impact of just one would neglect the impact of the other. Additionally, we would be unable to include both in the same regression models, as they are likely to be colinear.

One downside is the inability to explore the effects different types of absence (holidays during term time, illness, unauthorised absence) could have on academic outcomes. As a consequence, we are only able to estimate the average effect of overall absence. We therefore implicitly assume absence has the same effect regardless of the reason, which may or may not be the case. However, Klein et al (2024) found both types of absence harmed attainment equally, although they found *cumulative* unauthorised absences across schooling are more detrimental to attainment at the end of Key Stage 4 than cumulative authorised absences. We assume missing a day of school, when schools are open, will have the same effect on attainment regardless of the reason for absence.

There is a moderate negative relationship between "days absent" and "Attainment 8 scores" (correlation coefficient = -0.46).

Descriptive Statistics

Table 4: Descriptive Statistics by Cohort [18]

As can be seen in Table 4, there is only a small degree of variation in mean average absence and attainment between the academic cohorts used in this study. Attainment

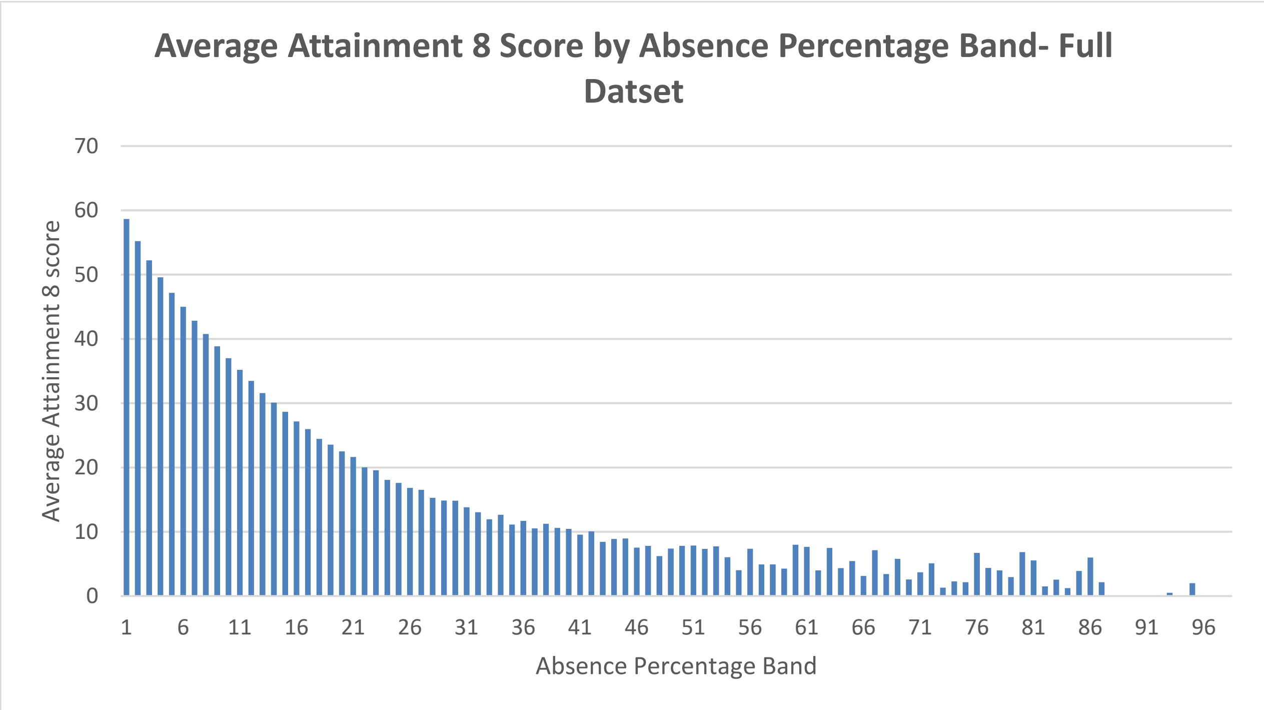
[18] Different government publications have different methodologies on how they report results and statistics, some of the figures in this report may differ from other official sources.

scores are standardised to make them comparable between cohorts. Combining the cohorts gives greater confidence when applying the results to future cohorts.

Table 4: Key Descriptive Statistics breakdown by cohort

Cohort	Number of Pupils	Mean ATT8 Score	Mean Maths GCSE Grade	Mean English GCSE Grade	Mean Absence Year 7-11
2016/17	466,624	47.2	4.6	4.7	44.3
2017/18	466,304	47.1	4.6	4.6	44.3
2018/19	485,747	47.2	4.6	4.6	45.6
Whole Sample	1,418,675	47.2	4.6	4.7	44.8

Figure 1: Average Attainment 8 score by Average Absence Percentage



Source: National Pupil Database

As Figure 1 shows, average Attainment 8 scores decrease as average absence increases. Those in the lowest absence percentage band (0-1%) had the highest (**58.7**) average Attainment 8 score. 0-1% absence equates to approximately 0 to 9 days of absence in 5 years. There are 149,096 pupils in this group, 10.5 % of the total sample. The effect becomes flatter after 40% absence.

Severely absent pupils, those with absence greater than 50%, have much worse academic outcomes than their peers. These pupils achieved an Attainment 8 score of

5.74 on average, compared to **47.3** for their peers. However, the number of pupils in this group is 2602, 0.2% of the sample.

Methodology

This report focuses on estimating the change in future earnings associated with marginal changes in school attendance. The analysis assumes that attainment acts as the mediating mechanism through which school absence affects earnings:

1 day of absence → Δ in human capital/skills accumulation → Δ in future earnings

In this model, Human Capital stock and skills accumulation is measured in **attainment** outcomes. The monetised estimate assumes attainment lies on the causal pathway between absence and earnings.

Our analytical approach involves two steps:

1. Estimating the association between absence and attainment: we analyse how changes in attendance influence academic achievement.
2. Translating attainment into lifetime earnings: we apply the established Department for Education (DfE) methodology[19] to monetise the effect of attainment on future earnings.

Non-linear relationship

A linear relationship would assume absence would have the same effect regardless of the prior levels of absence. For example, the 1st additional day of absence would affect attainment in the same way as the 101st day. Unlike the much of the academic literature (Gershenson, Jacknowitz, and Brannegan (2017); Liu, Lee, and Gershenson (2021); Cattan et al. (2017)) we do not assume a linear relationship between absence and attainment.

As can be seen from Figure 1 above, the relationship between Attainment 8 scores and absence doesn't appear to be linear. The trend shows a steady decrease in average Attainment 8 scores between 0% and 20% which then starts to curve before becoming flatter after 40% absence. We therefore test the assumption that there are diminishing marginal returns; the 101st day of absence will not have as great an effect as the 1st day. By conducting analysis using both linear and non-linear assumptions, we conclude that the effect of absence appears to be non-linear.

This finding could be considered intuitive: by a pupil's 100th day of absence, they will have already missed a large portion of schooling. Therefore, they won't have

[19] Lifetime Earnings Report: GCSE attainment and lifetime earnings

accumulated the skills required to achieve a high grade at GCSE and further absence will only have a small additional effect.

We model a non-linear (quadratic) relationship between school absence and attainment. We report the association between absence and attainment for the typical student, at the mean level (which a linear model would) and also those for who are persistently absent. This has policymaking value, by allowing us to use different values for the effect of absence on attainment for children at different stages of the absence distribution. We compared results from 3 models: a linear, a log and a quadratic approach. We concluded the quadratic approach was the best fit with the data. For further details see Annex E. To note, we do not report results for severely absent pupils. We also tested a regression approach where we separate pupils by absence quintiles and found evidence to support non-linear effects. Further details on this quintiles analysis can be found in Annex B.

As mentioned in the literature section, due to the gap between standardised testing in English schools – at ages 10 and 16 – it is difficult to establish if absence has a different relationship with attainment depending on the academic year. We therefore make the simplifying assumption that the effect is the same for each academic year in Key Stages 3 and 4.

Step 1: Effect of Absence on Attainment

As with the Aucejo (2016) study, we use an Ordinary Least Squares with fixed effects specification to estimate the relationship absence has with attainment, using the following equation:

$$\boldsymbol{y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 Z_i + \beta_4 c_i + \beta_5 p_i + \delta_{s_i} + \varepsilon_i}$$

Where $\boldsymbol{y_i}$ is the standardised attainment variable of student $\boldsymbol{i}$, $\boldsymbol{x_i}$ the number of days of absence between year 7 & 11, $\boldsymbol{Z_i}$ is an individual specific vector of background controls including SEN status, FSM provision, gender, major ethnic group and month of birth, $\boldsymbol{c_i}$ is cohort dummy, $\boldsymbol{p_i}$ is a control for previous attainment, $\boldsymbol{\delta_{si}}$ is a school fixed effect for attending school s , ε_i is an individual error term and α the constant (intercept).

As the effect is non-linear, we need to differentiate the above equation with respect to x to take the slope of the curve at different points. This gives the following:

$$\boldsymbol{\frac{d}{dx} = \beta_1 + 2 * \beta_2 x}$$

We need to include the prior number of days absence to calculate the slope: this is the *x* in the formula above. For our headline figure we use the mean number of days absent from the sample, where *x* = 44.8. We replace this number with 89 to calculate the effect size for persistently absent pupils.

As the dependent variable is standardised, the output is an effect size or percentage of a standard deviation. We then use this effect size in Step 2.

See Annex E for more details on how we selected our specification.

Step 2 Monetising the Impact

In step two we estimate the likely effect a change in school attainment will have on future lifetime earnings.

The lifetime earnings returns-to-attainment study by the Department for Education (DfE) estimated a one standard deviation increase in GCSE attainment is associated with a £115,000 (2024 prices) increase in present value lifetime earnings. There is nuance between "total GCSE points" and "Attainment 8 points", discussed in the data sources section. We assume the effect size would be the same for both and apply Attainment 8 effect size to the change in lifetime earnings associated with GCSE attainment.

Assuming the causal pathway holds, we can multiply the effect size generated in step one by the £115,000 figure above to arrive at our monetised figure.

$$\boldsymbol{E} = \boldsymbol{\beta} * \boldsymbol{le}$$

Where ***E*** = lifetimes earning loss, **β** is the effect size from step one and ***le*** is the lifetime earnings figure associated with a 1 SD change in GCSE scores.

We present the impact for the typical pupil with the average level of prior absence. We also provide estimates for persistently absent pupils. Due to the non-linear effects, we find the marginal effect of missing an additional day of school is greater for those with lower absence than those who are persistently absent. We do not provide policymaking recommendations based on these different results. We publish the marginal impacts for these pupils to help facilitate the cost benefit analysis of these policies, not to provide guidance on which groups to target for tackling absence.

Results

We use a quadratic specification to model the impact. These specifications can be interpreted as a 1 day increase in absence will result in a β x 1 SD decrease in Attainment 8 scores. However, the increase in absence will differ depending on the individual's prior attendance. Our headline figure uses the average number of days absence (44.8) for prior absence.

A persistently absent pupil will have been absent for a minimum of 89 days in total between Year 7 and Year 11. Severely absent pupils will have been absent for a minimum of 445 days, we do not report an effect size for this group. See Annex D for more details.

Attainment 8 Model Results

As attainment scores are standardised, the outputs from the modelling are presented as 'effect sizes'. The results are significant at the 0.1% level.

Dependent Variable: Standardised Attainment 8

Table 5: Results of Attainment 8 model

Coefficient β_1	Squared term Coefficient β_2	Adjusted R^2	Observations	Effect size at the mean
-0.0077 *** *CI [-0.0077, -0.0076]*	0.0000099 *** *CI [0.0000097, 0.000010]*	0.69	1,418,359	-0.0068 *(0.68% of a S.D)* *CI [-0.0067, -0.0069]*

Our model has an effect size of -0.0068 at the mean[20]. The results are statistically significant at the 0.1% level. The adjusted R^2 of 0.69 suggests the model can explain 69% of the variation in standardised Attainment 8 scores. This is relatively high for a model containing education outcomes. We also tested a linear model and this produced an effect size of 0.5% of a standard deviation, more closely in line with the literature. However, based on thorough testing we conclude that the relationship between absence and attainment is non-linear. Using our non-linear approach, the effect size at the mean is larger than in the linear model. For fuller details on how we established that non-linearity fit the data best, see Annex E.

To calculate the marginal impact of 1 *additional* day at different levels of prior attendance we use the formula:

$$\boldsymbol{\beta_1 + 2 * \beta_2 x}$$

Where β_1 is the coefficient on number of days absent, β_2 is the coefficient on square of days absent and x is the prior number of days absent.

Below we present the effect size for the mean level of absence, the median, and for the persistent (PA) and severely (SA) absent groups.

[20] This figure is not comparable to the figures in the literature review as it doesn't take it at the mean level of prior absence

Dependent Variable: Standardised Attainment 8

Monetised figures for different levels of prior absence (Attainment 8)

Table 6: Monetised figures for different levels of prior absence

Level	Prior No. Days Absent	Mean Average Attainment 8 points	Marginal effect size (% of a SD)	Monetised Low [21] estimate	Monetised Central estimate	Monetised High estimate
Mean x	44.8	49.5	-0.0068	£725	£772	£820
Median x	31.5	47.0	-0.0071	£757	£806	£856
PA x	89[22]	26.1	-0.0059	£629	£670	£711

Analysis of our dataset found that an additional day of absence, between Year 7 and 11, for the typical student with an average amount of prior absence was associated with a **0.68%** reduction in a standard deviation of Attainment 8 scores.

Monetising this effect, using the DfE lifetime earnings research, we estimate the reduction in future earnings, per additional day of absence, is **£772** (2024 prices). This falls to **£670** per day for PA students.

To deal with heteroscedasticity, we use HAC robust standard errors.

These estimated coefficients are in line with our starting assumption that there are diminishing marginal returns; the 101st day will *not* have as great an effect as the 1st day. The second derivative is positive, so the effect of absence diminishes asymptotically towards zero, as prior attainment increases.

There may be unobserved variables which may affect absence, attainment and earnings. Should these be included in the models above, the coefficient may decrease or become insignificant.

[21] The low and high estimates are calculated using the confidence intervals from the lifetimes earnings estimates, not those from we found in step one.

[22] This is the minimum number of days for persistent absence. The average number of days absent for those who were persistently absent but not severely absent was 141. See Annex D for more details

Risk and Uncertainty

The data, modelling and methodology has undergone thorough quality assurance by Department for Education (DfE) analysts. However, as previously stated there is only evidence to suggest a positive association between absence and attainment. Without a robust counterfactual to confirm attendance is causing a change in outcomes, we cannot claim causality therefore this evidence would be rated level **2 of 5 in the DfE evidence assessment framework**. As a result, we advise caution when using the findings. If these figures are used in cost benefit analysis, we recommend applying an appropriate level of optimism bias, to reflect this uncertainty.

Caveats

A monetised figure for one day of absence can be an important component in conducting cost benefit analysis. When using these estimates in analysis, we advise practitioners to highlight the following caveats.

Many factors will influence an individual's employment outcomes, and school absence is not the sole driver of attainment levels or future income. Pupils with high absence rates might have lower attainment and earnings than their peers due to other factors, regardless of the direct effect of absence.

Key Considerations:

- These are average effects, which should be applied to groups of pupils, rather than individuals. Earnings and attainment vary by individuals, even for those with identical levels of absence. There may be unobserved reasons for this; pupils with high absence rates may also possess unobserved traits such as lower motivation, poorer health, or strained relationships with school or teachers. Similarly, parental factors like education, occupation, and income can influence both attainment and earnings but are not fully captured in the analysis, potentially leading to omitted variable bias. Their exclusion may be affecting the findings, potentially inflating the monetised figures.

- The effect of an additional day of absence on attainment is non-linear, varying depending on how much schooling has already been missed. Our headline figures are for pupils with the average level of absence between Years 7 and 11.

- Our modelling suggests improving absence by 13 days is approximately associated with a 1 grade improvement at GCSE[23]. However, moving from a Grade 4 to a Grade 5 might impact future opportunities more significantly than

[23] 0.68% (effect size) x 11.2 (1 SD) x 13 (days) = 0.99 grades

moving from a Grade 1 to a Grade 2. This is explored in the returns to education publication (DfE, 2021). However, our analysis does not account for this nuance.

Broader Challenges in Analysis:

Estimating the impact of educational attainment and absence on labour market outcomes is inherently complex:

- **Labour Market Uncertainty**: The value employers attach to skills changes over time, and GCSE score distributions vary by cohort, leading to unpredictable effects on earnings.

- **COVID-19:** The pandemic disrupted GCSE exams and negatively impacted the labour market. We use pre-pandemic data which may not accurately reflect the post-pandemic world.

- **Structural Shifts**: Emerging trends like automation and artificial intelligence may cause longer-term shifts in employment, which are not accounted for in this report, which should be considered when appraising current policies, if based on these estimates.

Analytical Limitations:

- **No Experimental Counterfactual**: This is not an experimental evaluation with a counterfactual control group. We cannot know how the pupil would have performed if they had not had a period of absence.

- We have only included pupils with complete records, which may introduce an element of selection bias into the analysis.

- **Non-Causal Findings**: The results are not causal and should be interpreted as indicative trends within large samples rather than precise predictions at the individual level.

- Results at the extremes of the distribution: we find a quadratic model best fits the data for our dataset. However, this can produce results that are unintuitive at the extremes. For example, as Figure 3 shows, this approach implies that at extremely high absence levels, an additional day off school would be associated with positive economic returns. Given this appears unintuitive, we estimate marginal results at typical thresholds (mean absence, median absence and 10% absence), and do not report marginal results for pupils with severe absence.

Therefore we cannot apply these figures at an individual level, rather they are for inclusion in calculations when considering large samples of pupils.

- Analysis has been conducted on overall absence. We interpret these as effect of uncoordinated absence: individual pupil absence when the school was open. The results are therefore not directly applicable to coordinated absences (when the whole school is closed due to industrial action, extreme weather or COVID-19). However, we haven't stripped out coordinated absence due to data limitations. As uncoordinated is much more prevalent we use these results as a proxy for uncoordinated absence.
- Lifetime earnings represent a **private return to education** – they do not capture the effects of each person's education on the productivity, welfare, or labour market outcomes of others, or on the overall size of the economy.

Section Two - The Direct Relationship Between Absence and Labour Market Outcomes

The analysis in Section One of this report estimates the indirect effect of absence on lifetime earnings, by assuming school attainment as the mediation mechanism. This supplementary analysis aims to establish a *direct* relationship between school absence and earnings. Additionally, we use the same dataset to establish if there is a relationship between absenteeism and a range of labour market outcomes. While these findings shouldn't be used in cost benefit analysis calculations, they are important for setting context to policies and underpin the conversation around good attendance being a factor in building skills for use in the future workplace.

Data Source

The analysis in this section uses the Longitudinal Educational Outcomes (LEO) panel dataset. This dataset links the NPD data used in the main analysis, to earnings data from His Majesty's Revenues and Customs (HMRC) tax records and benefits records from the Department for Work and Pensions (DWP).

Analysis is conducted on a single cohort of pupils who attended state-funded schools and took their GCSEs during the 2006/07 academic year. It should be noted this is a different cohort of pupils to those in Section One.

There is a key limitation to this dataset; the latest age at which we can model the relationship between absence and earnings is 28. Evidence suggests that an individual's annual income peaks in 40-49 age bracket. Using data from the Annual Survey for Hours and Earnings (ASHE)[24] the House of Commons Library[25] suggests the average wage for 22-29 year olds was £621 per week compared to £823 per week for 40-49 year olds. This average includes those who attended independent schools and is therefore not comparable to the wage data our modelling.

Despite this, we felt it was still reasonable to measure the impact of absence on earnings at this age, given most individuals will be out of education and established in their careers by this age, and it provides an opportunity to assess the direct relationship between absence and labour market outcomes. We recommend that the figures from this section should not be used in cost benefit analysis calculations and policymakers instead use the lifetime earnings estimates from Section 1 to capture the full earnings impact.

[24] ONS ASHE: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2024

[25] House of Commons Library: https://commonslibrary.parliament.uk/research-briefings/cbp-8456/

Variables

In the first section we use control variables that might influence attainment whereas here we use those which might influence earnings. It should be noted that some of the characteristic variables may have different impacts depending on the dependent variable. For instance – female pupils tend to perform better in their GCSEs, however they have lower average earnings than males.

Where possible we use the same variables as the main analysis from Section One described in table 2. The definitions of FSM and SEN are largely the same in 2006 and 2016 and are used as dummy variables in both models.

Labour Market Variables

Earnings Data

In our dataset, total earnings can either be income from an employer or declared self-employed earnings. We include individuals who have zero earnings for the tax year in the modelling. Data was collected during the 2019/20 tax year when the individuals were 28 years old.

The 2019/20 tax year ended on the 5th of April 2020 and was unaffected by the furlough scheme which launched on 20th April. However, it should be noted that the first UK lockdown commenced on the 23rd of March 2020, 13 days before the tax year ended. This has the potential to have a minor impact on earnings data. We felt this was an acceptable risk and chose not to use the previous tax year when the individual would have been 27.

Similarly, we don't use earnings data from 2020/21 onwards due to the potential for the effects of the COVID-19 pandemic to affect the modelling.

Employment Status

We use two measures for employment status, **Sustained Employment** and **Sustained Benefits**. These are binary outcomes in our models.

Individuals who have 1 day of earnings in each of the 12 months of the 2019/20 tax year, are not on benefits for a sustained period, have greater than 100 days working will have a 1 for **Sustained Employment** [26], 0 otherwise. Individuals who claimed benefits for 1 day in 6 consecutive months of the 2019/20 tax year, and more than 10 days on benefits

[26] The definition of "sustained" is arbitrary and defined by DfE analysts.

and fewer than 100 days in employment will have a 1 for **Sustained Benefits**, 0 otherwise.

For clarity, one is not the inverse of the other. In cases where an individual is economically inactive, a student, or on a zero-hours contracts, these individuals would *neither* be in sustained employment nor in receipt of benefits.

Attendance

We have absence data for 2005/06 and 2006/07 for these individuals, when they were in Years 10 and 11. This differs from the indirect analysis (Section One) where we use absence across Years 7 to 11.

Persistently absent (PA) is defined as missing more than 10% of possible sessions (in line with DfE's definition[27]) during KS4. Unlike Section One, we model some employment outcomes for severely absent pupils, defined as missing 50% of possible sessions. These are binary variables in our dataset.

Descriptive Statistics

Our dataset is comprised of 637,790 individual pupils who took their GCSEs in 2006/07 academic year. Of those there are 511,227 with attendance records.

Table 7: Descriptive Statistics

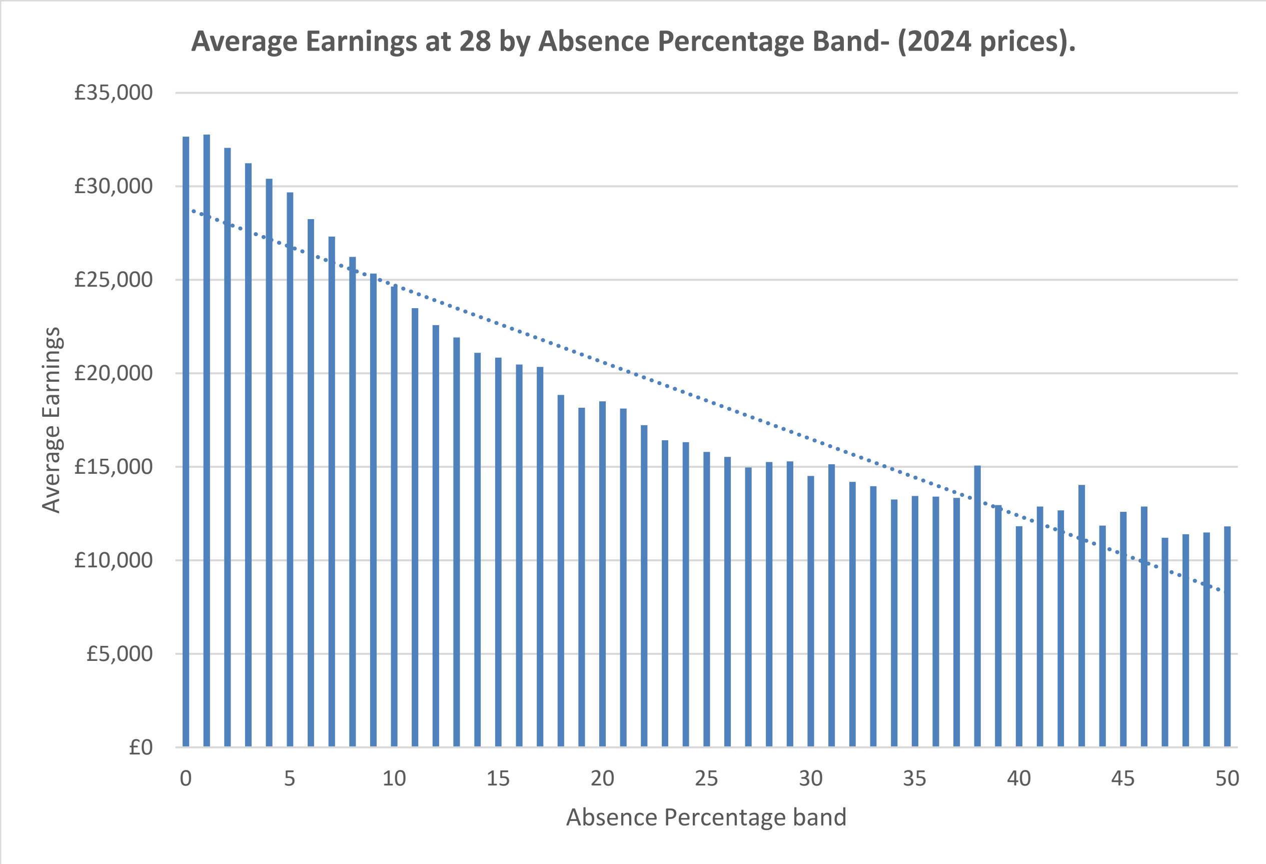
Mean Possible Days	Mean Days Absent Year 10-11	Mean Total Earnings (2024 prices)	Median Total Earnings (2024 prices)	Persistently Absent Pupils	Severely Absent Pupils
278	24.5	£26,426	£24,391	145,088	5389

Table 7 Descriptive Statistics for the LEO dataset used in direct analysis

Figure 2 shows the average earnings for each percentage band of absence in KS4.

[27] Attendance reporting methodology Pupil attendance in schools, Methodology - Explore education statistics - GOV.UK

Figure 2: Average Earnings at 28 by Absence Percentage Band (2024 prices)



Source: Longitudinal Educational Outcomes (LEO)

Figure 2 shows the earnings averages at 1 percentage point intervals in this dataset. It demonstrates those who have a higher percentage of absence during Key Stage 4, earn less on average than peers who have higher attendance.

The 33,629 pupils who were absent between 0-1% between Years 10 and 11 earned, on average, approximately £33,000 in 2024 prices. The 18,303 pupils who missed 10-11% of their possible sessions earned approximately £25,000, in 2024 prices, on average. Those who missed between 50-51% of their possible sessions earned approximately £12,000, in 2024 prices, on average. These figures do not control for any socio-economic characteristics which may affect earnings.

To note, earnings are constrained by the fact that the UK has a minimum wage which acts as a lower bound for fulltime workers. As individuals can't earn negative values, 0 is the lower bound for part time and unemployed individuals. We are unable to establish if an individual is full or part time as the LEO dataset does not include hours worked.

Average Earnings by Absence Band

Table 8: Average Earnings by Absence Band

Absence Band	Number of Pupils (n)	Average earnings at age 28 (2024 prices)
0-1%	33,530	£32,200
0-9.9%	366,139	£29,800
10-100% Persistently absent	145,088	£19,500
50-100% Severely absent	5389	£10,100

As table 8 shows, persistently absent pupils earned more than £10,000 less on average (2024 prices) than those who were absence less than 10% of possible sessions. This rises to almost £20,000 less for those who are severely absent. These figures do not control for any socio-economic characteristics which may affect earnings.

Methodology

To model the association between absence and wages at 28 we use a similar specification to the analysis in the first section.

Absence isn't normally distributed; it is right skewed. In this dataset, 72% of pupils (366,139) had absence between 0% and 9.9%. 28% of the pupils (145,088) had absence between 10% and 100%. Only 1% of the dataset is made up of severely absent pupils (5389), who missed more than 50% of possible sessions.

We assume a quadratic relationship between absence and earnings, so we include absence squared (as we did in Section One).

Earnings data at age 28 is not normally distributed; we deal with this by log transforming the earnings data.

We use a logistic regression for the labour market outcomes modelling as the dependent variables are binary.

Specification 1: Impact of Key Stage 4 absence on Earnings at Age 28

Similar to the analysis in Section One, we use an Ordinary Least Squares with fixed effects specification to estimate the relationship absence has with earnings.

We employ a log-linear specification using the following equation:

$$\boldsymbol{ln\ y_i + 1 = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 Z_i + \beta_4 c_i + \beta_5 p_i + \delta_{s_i} + \varepsilon_i}$$

Where y_i is the natural logarithm of earnings at 28 for individual i, x_i is the number of days of absence/absence percentage between year 10 & 11, Zi is an individual specific vector of background controls including SEN status, FSM provision, gender, ethnic group , p_i is a control for previous attainment, δs_i is a school fixed effect for attending school s, ε_i is an individual error term and α the intercept.

As the effect is non-linear, we need to differentiate the above equation with respect to x to take the slope of the curve at different points. This gives the following:

$$\boldsymbol{\frac{d}{dx} = \beta_1 + 2 * \beta_2 x}$$

We need to include the prior number of days absence to calculate the slope, this is the x in the formula above. For the headline estimates we calculate this at the mean level of absence.

Specification 2: Impact of Key Stage 4 absence on Labour Market Outcomes

We use a logistic regression to estimate the odds of those who are either persistently or severely absent of obtaining different employment status.

$$\boldsymbol{Logit(Yb, Ys)_i = \left(\frac{P_i}{1-P_i}\right) = \beta_1(a_i) + \beta_2 X_i + \varepsilon_i}$$

Where $\mathbf{Y_b}$ is the logistic function of receiving benefits for 6 months at 28, $\mathbf{Y_s}$ is the logistic function of being in sustained employment for 12 months at 28, $\mathbf{a_i}$ is a binary variable to indicate a pupil who has missed either **>10%** or **>50%** of sessions, $\mathbf{X_i}$ is a vector of characteristic variables, and $\boldsymbol{\varepsilon_i}$ is the error term.

The outputs of logistic regressions represent the natural logarithm of the odds. To obtain the odds, we exponentiate these outputs. An exponentiated coefficient of 1.5 can be interpreted as a one-unit increase in the predictor increases the odds of the outcome by the 50%. A coefficient of 1 would mean the odds of the event occurring are the same as not occurring.

Results

Results of Specification 1: Impact of Key Stage 4 absence on Earnings at Age 28

Table 9: OLS Model Results

Independent Variable	Coefficient β_1	Squared term Coefficient β_2	Adjusted R^2	Observations	Effect at the mean
Days	-0.00871 *** CI [-0.008, -0.009]	0.000026 *** CI [0.00002 , 0.00003]	0.13	466,968	**-0.008** CI [-0.007, -0.008]

When controlling for factors known to influence wages, there is a statically significant (at the 0.1% level) relationship between days absent/absence percentage during KS4 and earnings at 28.

As the model specification is log-linear, the "days" coefficients can be interpreted as a percentage change in earnings rather than an absolute change.

For those with non-zero earnings at age 28, the mean level of absence is 22 days. At this level of absence, the effect of –0.008 means that for each additional day of absence during Key Stage 4, annual earnings at age 28 are expected to decrease by approximately 0.8%. As the mean average salary[28], this represents a decrease of approximately £150 (in present values) per day of absence. This figure does not include those with earnings equal to zero[29].

There may be unobserved variables which may affect absence, attainment and earnings. Should these be included in the regression, the coefficient may decrease or become insignificant.

[28] For those with positive earnings, this is £29,000 (2024 prices).
[29] To note, these 'direct' estimates aren't directly comparable with the 'indirect' estimates from Section 1. The indirect estimates are estimates of changes in lifetime earnings, accounting for individuals' earnings trajectories. The direct estimates here are static values at age 28, and only use attendance in Years 10-11, not the full range of Years 7-11. To account for present values, we discount using the HMT Green Book discount rate.

The figures from the direct analysis should **not** be used in cost benefit analysis calculations, as they are a snapshot of earnings from a single tax year. The analysis from the indirect analysis takes the trajectory of an individual's earnings into account.

The adjusted R^2 is 0.13, suggesting that the model can only explain 13% of the variation in earnings at age 28. A smaller R^2 is not unexpected in this context as socio-economic data typically contains a lot of unobserved heterogeneity. Factors like education, employment, and earnings are influenced by psychological, social, and institutional factors that are hard to fully capture in a regression model.

To deal with heteroscedasticity, we use HAC robust standard errors.

Specification 2a Logistic regressions sustained period (6 months) on benefits.

Table 10: Results of the sustained benefits logistic regressions

Dependent Variable	Coefficient (natural Log of Odds)	Odds (exponentiated coefficient)	Observations
Days Absence Continuous variable	0.016 ***	1.02	479,058
Persistent Absent (>10% Absence)	0.99 ***	2.7	479,058
Severe Absent (>50% Absence)	1.43 ***	4.2	479,058

Absenteeism is significantly linked with claiming benefits for more than 6 months at age 28. After controlling for socio-demographic factors known to influence employment outcomes, we find for each additional marginal day absent in Years 10 and 11, the odds of being on sustained benefits at age 28 increases by 2%.

Persistently absent pupils (>10% absence) are 2.7x more likely to be in receipt of benefits for a sustained period at 28 compared to their peers. This rises to 4.2x as likely for those who are classified as Severely Absent (>50% absence).

To note, 50,358 records were not included in this model due to missing benefits data. The number of observations is 479,058.

Specification 2b Logistic regression sustained period (12 months) of employment.

Table 11: Results of the sustained employment logistic regressions

Dependent Variable	Coefficient (natural Log of Odds)	Odds (exponentiated coefficient)	Observations
Days Absence Continuous variable	-0.014	0.99	479,058
Persistent Absent (>10% Absence)	-0.86	0.42	479,058
Severe Absent (>50% Absence)	-1.36	0.26	479,058

Absenteeism is significantly negatively associated with being in sustained employment (12 months) at age 28. After controlling for socio-demographic factors known to influence employment outcomes, we find as the number of days absence increases, the odds of being in sustained employment at 28 decreases by 1%.

Persistently absent pupils (>10% absence) are 58% less likely to be in employment for a sustained period at 28 compared to their peers. This drops to 74% as likely for those who are classified as Severely Absent (>50% absence).

To note, 50,358 records were not included in this model due to missing benefits data. The number of observations is 479,058.

Caveats

The relationship between absence and future labour market outcomes can be important for setting any context for discussions around the importance of good attendance. However, we advise on the caveats below being included in any analysis which uses the figures from within this section of the report.

Many factors will influence an individual's employment outcomes; school absence is not the sole driver of future income or employment prospects. Pupils with high absence rates might have lower earnings than their peers due to other factors, regardless of the direct effect of absence.

Key Considerations:

- School absence explains a greater variation in school attainment than labour market outcomes. Our attainment models in Section 1 have a good fit, explaining approximately 70% of the variation in GCSE results, which can be considered high. However, this falls to around 18% for the direct earnings models in Section 2. This is to be expected, as educational outcomes are only one factor affecting individual earnings.
- There are many more unobserved reasons for this unexplained variation in earnings returns. This could include personal motivation, health, personality traits and more. Similarly, parental factors like education, occupation, and income can influence earnings but are not fully captured in the analysis, potentially leading to omitted variable bias. Their exclusion may be affecting the findings, potentially inflating the monetised figures.
- The effect of each additional day of absence is non-linear, but our headline figures are for pupils with the average level of absence during Key Stage 4. For the reasons listed above we report average effects, the results of which should not be applied to individuals.
- The indirect analysis in Section 1 uses more years of educational data than the direct analysis in Section 2, by including Years 7-9. Additionally, the direct analysis focuses on a different cohort to the those in the main analysis and are therefore not directly comparable.
- As with the direct analysis in Section 1, this analysis does not distinguish between the impacts of absences for different reasons. We do not estimate differential impacts of authorised and unauthorised absence, for example. We interpret our estimates as relating to the effect of uncoordinated absences.

Analytical Limitations:

- No Experimental Counterfactual: This is not an experimental evaluation with a counterfactual control group. We cannot know how the pupil would have performed if they had not had a period of absence.
- We have only included pupils with complete records, which may introduce an element of selection bias into the analysis.
- Non-Causal Findings: The results are not causal and should be interpreted as indicative trends within large samples rather than precise predictions at the individual level.

Finally, the main caveat here is that this is a snapshot of one tax year for an individual. We have no details of any circumstances which surround their employment status at age 28. These circumstances may change by age 29 for example. Additionally, wages do not peak at this age. Different individuals may follow different paths to their optimal level of earnings which may not be reflected in their earnings in their late 20s. We therefore strongly advise against using these findings in cost benefit analysis.

Overall Conclusion

The main analysis points to a statistically significant (at the 0.1% level) negative association between school absence and attainment, consistent with the literature. As overall absence increases, attainment outcomes at the end of KS4 decrease, when controlling for other factors known to influence achievement.

As previous research points to a link between attainment and income we assume attainment indirectly acts as a mediation mechanism between absence and earnings. Using our modelling findings we estimate a monetised figure for one day of absence which can be used in the economic appraisal of policies which aim to tackle high levels of absence.

We use earnings data to test if there is a direct association with earnings. We find a statistically significant link (at the 0.1% level) between school absence and a range of labour market outcomes at age 28. As absence increases, earnings decrease, when controlling for other factors known to influence income. Additionally, we find increased absence was associated with a higher chance of being on benefits for a sustained period, and a lower chance of being in sustained employment at the age of 28.

We report the average effect of absence for all pupils in state-funded secondary-schools. We do not report breakdowns by different groups of students by different sociodemographic backgrounds.

Caution is advised when interpreting or using these estimates. Any analysis that uses these estimates should acknowledge the observed associations between absence and attainment and attainment and earnings may not be accurate predictions of how absence may affect future cohorts. Any figures used in any modelling should be accompanied by the caveats listed within this report.

Further Research

There are further areas of interest which are out of scope in this report.

- One such area might be to establish if missing a day earlier in the school life has the same effect as missing a day as the pupil progresses. If pupils in primary school are absent, then we might reasonably assume there is a fade-out effect by the time they take their GCSEs in Year 11 (Bailey et al., 2020). Conversely, there is an argument that '*skills-beget-skills'* (Cunha and Heckman, 2007). A shock to attainment may have a scarring effect, forcing pupils onto a lower trajectory of human capital accumulation with the effects magnify over time.

- Another area of interest is coordinated absence. Absence data is now collected daily from schools, making future research projects involving the effects of snow days or strikes easier to undertake.
- It may be feasible to include the Instrumental Variables used in the Aucejo and Romano (2016), Gaete-Romeo (2018), and Goodman (2014) studies in future studies. Since 2022, the date of absence has been collected, which could help point to a causal relationship between absence and attainment using the department's data.
- We model how an increase in absence can have a negative relationship with attainment. We currently assume the inverse is true, improving attendance will have a positive relationship with attainment and earnings. Further work could be carried out to establish if this assumption holds.
- Finally, this report uses pre-pandemic data. We suggest revisiting this research to understand if the effect of absence on attainment has changed since the COVID-19 pandemic. The earliest this could be completed, without pandemic-related school closures affecting attendance records, is for those sitting their GCSEs in 2025/26.

References

Aucejo, E. M., and T. F. Romano. (2016). “Assessing the Effect of School Days and Absences on Test Score Performance.” *Economics of Education Review* 55: 70–87. https://doi.org/10.1016/j.econedurev.2016.08.007.

Bailey et al (2020) Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions (Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions - Drew H. Bailey, Greg J. Duncan, Flávio Cunha, Barbara R. Foorman, David S. Yeager, 2020 (sagepub.com))

Cattan, S., D. A. Kamhöfer, M. Karlsson, and T. Nilsson. (2022). “The Short- and Long-Term Effects of Student Absence: Evidence from Sweden.” Discussion Paper No. 10995. Bonn, Germany: IZA. https://academic.oup.com/ej/article/133/650/888/6825447

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)

(Cunha and Heckman, 2007) The Technology of Skill Formation (The Technology of Skill Formation - American Economic Association (aeaweb.org))

Dräger, J., Klein, M., Sosu, E., (2022) “The long-term consequences of early school absences for educational attainment and labour market outcomes” 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

Gaete-Romeo, G. 2018. “Follow the Leader: Student Strikes, School Absenteeism and Persistent Consequences on Educational Outcomes.” https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2988825.

Gershenson, S., A. Jacknowitz, and A. Brannegan. 2017. “Are Student Absences Worth the Worry in U.S. Primary Schools?” *Education Finance and Policy* 12 (2): 137–65. https://doi.org/10.1162/EDFP_a_00207.

Gershenson, S., J. R. McBean, and L. Tran. 2018. “Quantile Regression Estimates of the Effect of Student Absences on Academic Achievement.” Discussion Paper No. 11912. Bonn, Germany: IZA.

Goodman, J. 2014. “Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time.” NBER Working Paper No. 20221. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w20221.

Gottfried, M. A. (2011).
“The Detrimental Effects of Missing School: Evidence from Urban Siblings.” *American Journal of Education* 117 (2): 147–82. https://doi.org/10.1086/657886.

Gottfried, M. A., and J. J. Kirksey.
(2017). “‘When’ Students Miss School: The Role of Timing of Absenteeism on Students’ Test Performance.” *Educational Researcher* 46 (3): 119–30. https://doi.org/10.3102/0013189X17703945.

Hodge, L., A. Little, and M. Weldon. (2021). “GCSE Attainment and Lifetime Earnings.” Research Report. Department for Education (2021). https://assets.publishing.service.gov.uk/media/60c36f0cd3bf7f4bd11a2326/GCSE_Attainment_and_Lifetime_Earnings_PDF3A.pdf

Liu, J., M. Lee, and S. Gershenson. (2021). “The Short- and Long-Run Impacts of Secondary School Absences.” *Journal of Public Economics* 199: 104–441. https://doi.org/10.1016/j.jpubeco.2021.104441.

Sims (2020). “Briefing note School absences and pupil achievement”, available at (https://repec-cepeo.ucl.ac.uk/cepeob/cepeobn1.pdf) accessed 17/02/2025

Annex

Annex A

Regressions by Cohort

To test if the effect was the same for each academic year we ran the main analysis in for each of the cohorts separately. As can be seen in table 15, the effect size at the mean doesn't undergo a noteworthy change between each cohort.

Table 12: Results by cohort split

Academic Year	2016/17	2017/18	2018/19	Full Dataset
Average number of days absent Y7 - 11	44.3	44.3	45.6	44.8
Observations	466,308	466,304	485,747	1,418,359
Effect size (% of SD)	-0.0068	-0.0069	-0.0069	-0.0068
Adjusted R squared	0.69	0.69	0.69	0.69
Monetised Marginal Effect	£772	£783	£783	£772

We therefore have confidence in using these results to predict future impacts from absence.

Annex B

Quintile Regressions

The findings from our analysis suggest that an additional day of absence will have a different effect on attainment depending on the pupil's prior absence. A pupil who has missed a significant portion of school is, on average, already unlikely to perform well at GCSE. Therefore, a marginal day of absence is unlikely to have a large effect on results

We test this by grouping the data by absence quintile and running an OLS with fixed effects regression on each split. Those in Quintile 1 have the lowest levels of absence, and Quintile 5 have the highest. If the effect was linear, we would expect the effect size from each model to be similar.

Absence Quintile Regression Outputs

Table 13: Results by absence quintile.

Attendance Quintile	1	2	3	4	5
Average number of days absent Y7 - 11	8.0	19.4	31.7	49.8	114.9
Observations	283,672	283,672	283,672	283,672	283,671
Effect size (% of SD)	-0.014 *** *CI [-0.001, -0.018]*	-0.0087 *** *CI [-0.0078, -0.03]*	-0.0072 *** *CI [-0.001, -0.02]*	-0.0062 *** *CI [-0.0034, -0.016]*	-0.0049 *** *CI [-0.0047, -0.0051]*
R squared	0.65	0.64	0.62	0.61	0.61
Monetised	£1590	£988	£817	£704	£556

The table above displays a sizable difference between the monetised effect sizes between the 1st and 5th quintiles. Therefore, we have confidence in stating the effect in not linear. As there is unlikely to be demand for multiple effect sizes, we report the effect size at the mean absence average (approximately 45 days).

However, some policies may specifically target persistently absent pupils (those missing at least 10% of schooling). The average effect figure won't be appropriate to use in these cost benefit calculations. In addition, we publish figures for these sub-categories alongside the average effect. See results section from Section 1.

Annex C

Full or Filtered data

The dataset used for the analysis in Section 1 includes certain records which have the potential to adversely affect the outputs.

Pupils who attend Special Educational Needs schools or alternative provisions may have good attendance, but their circumstances may prevent them from achieving good GCSE results. There were approximately 2% of the pupils (n=29,069) pupils in the data who attended these settings. We define pupils who didn't attend these settings as "mainstream"

Approximately 16% of the pupils (n= 230,000) had an improbable number of *potential* sessions. This issue often arises when a pupil changes schools during the academic year, resulting in duplicate records. While we take care to remove all duplicate entries, the remaining record may only cover a short period of recorded absence. Additionally, human error during the inputting of attendance data into school management systems may further distort the findings. We defined students who attended for a "full year" as having +- 10% of the mode average number of potential sessions for the academic year by cohort.

In our dataset, 0.2% of the pupils (n = 2645) were classed as severely absent. This group of pupils miss more than 50% of possible sessions, which amounts to over 190 sessions or 95 days per academic year. On average, severely absent pupils experience worse attainment outcomes than those who attend more than 50% of the school year. The average Attainment 8 score for severely absent pupils is **5.7** compared to **47.4** for non-severely absent pupils. However, there are outliers whose attainment has the potential to add noise to the analysis; one severely absent pupil had an Attainment 8 score of **90**. In comparison, the highest score for the non-severely absent pupils was **95.25**.

To test the theory that these pupils would adversely affect the modelling, we split the data into 4 different sub samples as described below:

Table 14: Descriptive statistics by filtered dataset

Sample	n (pupils)	Correlation absence and attainment	Mean average days absent Year 7-11	Standard Dev Days absent	Mean Attainment 8 score	Standard Dev Attainment 8 score
1. Full Dataset	1,418,675	-0.455	44.8	48.4	47.2	19.4
2. Filtered to Mainstream	1,389,293	-0.415	42.3	43.1	48.0	18.7
3. Mainstream and "full year" (+- 10% of the mode)	1,328,368	-0.411	41.8	42.4	48.2	18.6
4. Mainstream, "full year" with no Severe Absence (>50%) pupils	1,327,243	-0.414	41.4	40.3	48.3	18.6

We ran models with a restricted dataset which removed the above-mentioned pupils, and unrestricted where we included all pupils with complete data – the full dataset.

Model Outputs by different filtered datasets

Table 15: Model Outputs by different filtered datasets

Data sample	Full data set	Mainstream Only	Full year	No Severely Absent pupils
Coefficient	-0.0068 *** *CI [-0.0067, -0.0069]*	-0.0067 *** *CI [-0.0067, -0.0068]*	-0.0069 *** *CI [-0.0069, -0.0068]*	-0.0071 *** *CI [-0.0071, -0.0072]*
Adjusted R^2	0.68	0.66	0.66	0.66
Observations	1,418,359	1,389,293	1,328,368	1,327,243

As can be seen in table 15, we found restricting the data did not result in a noteworthy difference between correlations, model coefficients or adjusted R^2. We therefore used the full dataset to account for as many pupils as possible.

Annex D

Total Days Schooling

The total number of days in secondary schooling will differ for each pupil for a range of reasons. All state-funded schools are required to provide two possible sessions per day: one in the morning and one in the afternoon. Schools must be open to all pupils for *at least* 380 sessions or 190 days during any school year, although some schools are open more.

We only count absence during autumn and spring term in Year 11. This is due to pupils starting their GCSEs in the summer term and going on study leave.

Summer Term starts on a different date each year – depending on when the easter holidays end. Easter Sunday fell on the following dates:

- **2016/17:** 16th April 2017
- **2017/18:** 1st April 2018
- **2018/19:** 21st April 2019

Additionally, Local Authorities and some schools can set different dates for their school holidays.

We assume a school year is 190 days for years 7, 8, 9, and 10. We assume the school year is approximately 130 days in year 11. This gives a total number of days spent in secondary school as approximately **890** days (*190 x 4 + 130*).

Based on this figure we estimate a pupil who is persistently absent will miss a minimum of **89 days** (890 x 10%) and a severely absent pupil will miss **445 days** (890 x 50%).

Annex E

Specifications

For the analysis in Section 1 estimating the effect of school absence on attainment, we tested three specifications before settling on our preferred identification approach.

We initially started with a linear model, in line with most of the economic literature on school absence. Following discussions with analytical colleagues, we then explored non-linear models, using log-transformed and quadratic models.

Specification 1: **Linear** $\boldsymbol{Y = \beta_0 + \beta_1 x + \epsilon}$

Specification 2: **Log-Transformed** Independent Variable $\boldsymbol{Y = \beta_0 + \beta_1 ln(x) + \epsilon}$

Specification 3- **Quadratic** Model $\boldsymbol{Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon}$

Testing Different Specifications

To choose the most appropriate model, we ran a series of econometric tests and consulted with a range of experts across and beyond the department. Econometric tests included: a Breaush-Pagan test for heteroscedasticity; an F/Wald test and AIC/BIC tests. The results are presented in Table 16 below.

We decided to pursue Model 3 (quadratic) based on the results of these econometric tests. Model 3 (quadratic) has the lowest AIC (2,370,174) and BIC (2,433,785)**,** meaning it is the best model among the three in terms of goodness of fit while accounting for complexity. This is then followed by Model 1 (linear) and then Model 2 (log) respectively. We also use the quadratic approach over the log approach given concerns around using a log transformation on school absence – a discrete variable. We use HAC standard errors to counter issues of heteroskedasticity.

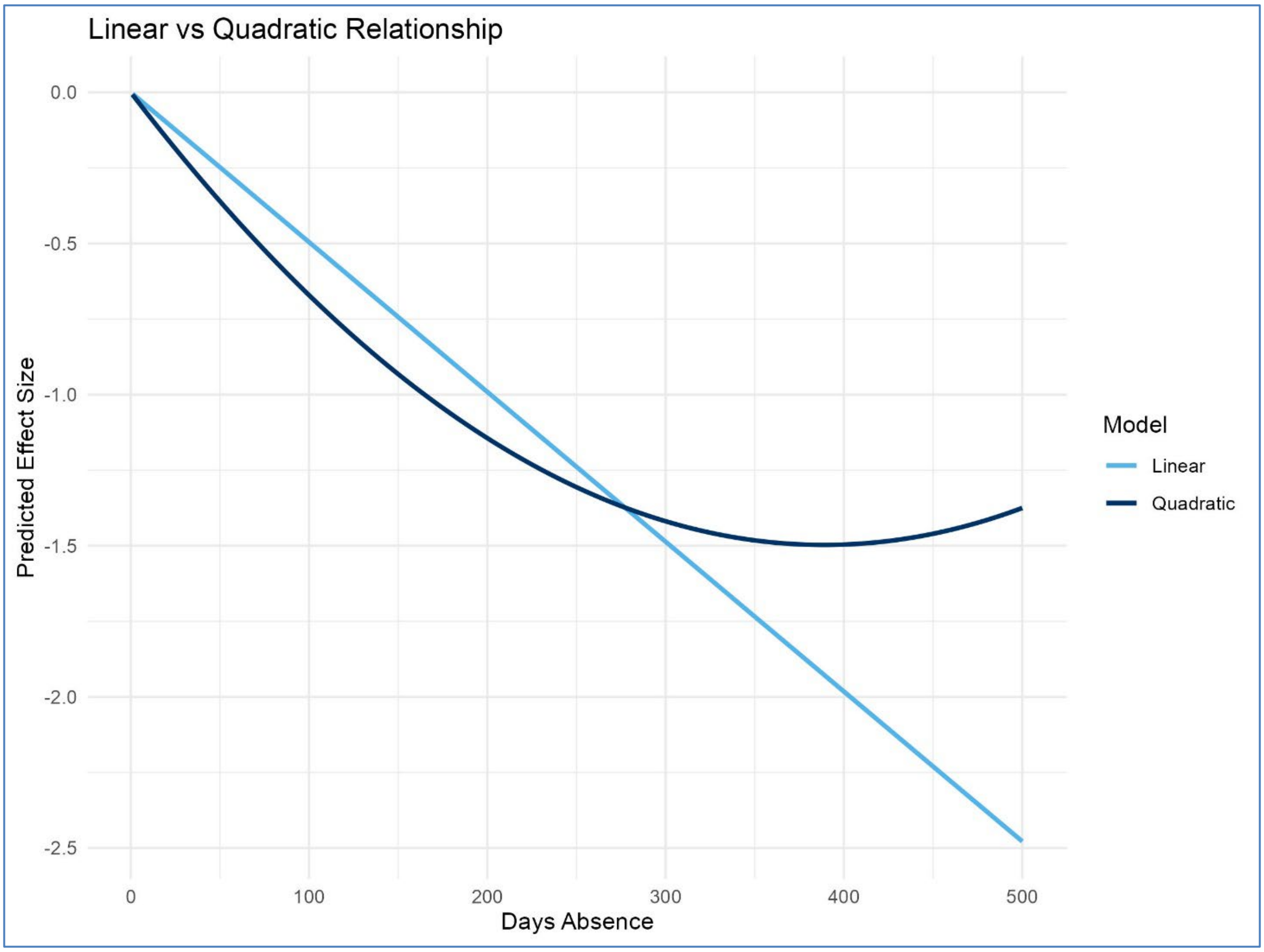
We find that effect sizes vary somewhat by model, although all are of a similar magnitude. At the mean, the quadratic model has the largest effect (0.68% of SD), followed by the log model (0.55% of a SD) and the linear (0.50% of a SD). All values fall within the broad economics literature of the effect of absence on attainment.

Table 16: Tests on 3 different specifications.

	Linear	Log-Transformed	Quadratic
Coefficient	-0.004956***	-0.245478***	-0.00769516***
Effect size		-0.00548 at the mean	-0.00681 at the mean
Adj R2	0.68423	0.68319	0.68975
AIC	2395190	2399838	2370174
BIC	2458788	2463436	2433785
Wald test	Stat: 23,085.1 P-val: < 2.2e-16 Conc: coefficient not 0		Stat: 32,800.3 P-val: < 2.2e-16 Conc: coefficient not 0
Breusch Pagan	BP Stat: 160.48 DoF: 118 P-val: 0.005681 Conc: heteroskedastic	BP Stat: 138.96 DoF: 118 P-val: 0.09114 Conc: *not* heteroskedastic	BP Stat: 8083.6 DoF: 118 P-val: 0.00000000000000022 Conc: heteroskedastic
F test		ChiSq: 227709953 P-val: 0.00000000000000022	

Figure 3 below demonstrates how the linear and quadratic approaches vary in effect size throughout the absence distribution. Given the nature of a quadratic model, results become unusual at extremely high or low values of absence. For example, students missing over 500 days of school between Years 7 and 11 would are assumed to have positive returns to missing an extra day of school under this specification, a finding that is unintuitive. As such, we focus the results of the quadratic model on the majority of students in the absence distribution, particularly at key absence thresholds – the mean, median and those considered 'persistently absent', or missing 10% of school (around 89 days across the 5 years).

Figure 3: Plot of linear and quadratic relationship



We conclude that Model 3 (the quadratic approach) is the most appropriate model, based on the results of our econometric tests. This is supported by the results of the AIC and BIC tests, suggesting it strikes the best balance between fit and complexity. We use this for our headline figures. We caution against using the results for pupils with extremely high levels of absence.

Annex F

Association with Maths and English

We replaced standardised Attainment 8 with Attainment 8 points, Maths GCSE Grade and English GCSE Grade

Table 17 Effect sizes for Maths and English

Variable	Coefficient β_1	Coefficient β_2	Marginal effect at mean	Adjusted R^2
Attainment 8 points	-0.149 ***	0.0002 ***	-0.13	0.69
Standardised Attainment 8	-0.008 ***	0.000010 ***	-0.0068	0.69
Maths GCSE Grade	-0.014 ***	0.00002 ***	-0.012	0.68
Standardised Maths	-0.007 ***	0.000009 ***	-0.0059	0.68
English GCSE Grade	-0.010 ***	0.00001 ***	-0.010	0.53
Standardised English	-0.005 ***	0.000006 ***	-0.0048	0.53

Table 17: Outputs for different attainment variables as the dependent variable

Our individual subject results are in line with the previous research (Gershenson, Jacknowitz, and Brannegan (2017) - 0.2-0.7 (Maths), 0.2-0.4 (Reading). Like the literature, absence is associated with a greater impact on Maths results than English. One potential explanation for this finding is that reading is relatively easier to access at home compared to Maths, although this is not based on empirical evidence.

The adjusted R^2 for English also suggests our model explains less variation in the English grade than Maths.

Department for Education

Reference: RR or RB number R1498

ISBN: 978-1-83870-625-8

For any enquiries regarding this publication, contact us at: Team.CentralEconomics@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications