Absence and attainment: Evidence from pandemic policy

Stephen Gibbons, Sandra McNally and Piero Montebruno

September 2024

Abstract
A high level of school absence has persisted across many countries since the COVID-19 pandemic. We use English data to investigate whether a student's absence during the pandemic had a causal impact on school attendance and academic progress in future years, using variation in local regulations during the pandemic (not aimed at schools). We find that more stringent regulations caused higher rates of school absence at that time, leading to lower attendance and rates of achievement in subsequent years. Our evidence suggests that the persistent effect is caused by changes in parents' and pupils' attitudes to attendance and not because of rules forcing students to stay at home when they had been in contact with others who had COVID-19. The effects of policy restrictions on contemporaneous and persistent absences was stronger for lower socio-economic groups.

Keywords: student absence, COVID-19 pandemic
JEL: I24; I28

Funding provided by the Nuffield Foundation grant number FR-000022949. We are grateful to the Nuffield Foundation for funding this research and to members of the advisory group set up to support this project. This work was undertaken in the Office for National Statistics Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners.

Stephen Gibbons, London School of Economics and Centre for Economic Performance at London School of Economics. Sandra McNally, University of Surrey and Centre for Economic Performance at London School of Economics. Piero Montebruno, Centre for Economic Performance at London School of Economics.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

1. Introduction

Globally, children experienced long periods of exclusion from school during the COVID-19 pandemic. Since then, absence rates have remained very high in many countries, with huge increases in 'chronic' absenteeism where students are regularly missing from school on a weekly basis (Mervosh and Paris 2024, Adams 2024, Taylor 2023). These high levels of absenteeism are of obvious concern, given the potential impact on educational outcomes and inequalities, and attendant problems like crime and poor labour market engagement. Many emerging studies worldwide have documented higher rates of absence and a drop in educational achievement and progress post-pandemic. There has been speculation that the tolerance of absence or enforcement of home-schooling during the pandemic gave rise to a new culture of persistent absenteeism in its aftermath, but qualitative evidence suggests points more towards COVID-related anxiety as the aggravating factor (McDonald, Lester and Michelson 2022). But these inferences are mostly based on before-after comparisons of absence and achievement or small qualitative surveys, rather than causal analysis. Recent studies for the US have looked at how geographical variation in opportunities for, and student uptake of, in-person teaching during the pandemic affected student outcomes (Dee 2024, Jack et al 2023, Goldhaber et al 2023, Ross et al (2024). No research has yet looked directly at whether a student's absence during the pandemic had a causal impact on attendance and academic progress in future years.

This study aims to answer the question of whether a student's absence during the pandemic affected their post-pandemic attendance and academic achievement. We find that absence induced by health and social policies that closed businesses and restricted social contact during the early stages of the pandemic – autumn 2020 - caused higher rates of school absence at that time, leading to lower attendance and lower rates of achievement in subsequent years (2021/22). Our focus on autumn 2020 is motivated by the fact that, at this time, a regime of local regulations was in place (the Tier Regulations) that generated differences in the social and economic environment in which different schools were operating. These policies lead to variation in absence rates across schools. Our evidence suggests this was caused by changes in parents' and pupils' attitudes to attendance. We find no effect from compulsory absence caused by rules that forced students to stay home when they been in contact with others who had recorded cases of COVID-19.

These conclusions are reached by comparing absence and academic progress of students in the pre-pandemic period 2017/18-2018/19 with the pandemic period 2020/21-2021/22. The research design is partly determined by data availability. Our data source offers high volume and detailed annual pupil-level administrative data for the whole of England, but was not released for the 2019/20 academic year, and there were no standard academic tests in 2019/20 or 2020/21. We therefore consider how absence in autumn of the 2020/21 year affected outcomes a year later in 2021/22. We use a pupil and/or school fixed effects design to control for the baseline relationship between absence in 2017/18 and outcomes in 2018/19. Given there are still potential unobserved pupil level confounders, we use a Two Stage Least Squares design in which pupil absence in autumn 2020 is predicted from policy variables measuring a school's exposure to the local COVID-19 restrictions. In this way, we estimate our relationships of interest from policy-induced quasi-experimental variation in absence rates in autumn 2020.

Our work contributes to three strands of literature: on the causes of student absenteeism; on the effects of absenteeism and lost education time on educational outcomes; and on the impacts of the COVID-19 pandemic on educational outcomes. We provide some unique innovations. Our paper is the first of which we are aware to look at the dynamics of absenteeism, that is the extent to which individual absenteeism persists from one year to the next. Secondly, we provide some of the first direct evidence that public health policies which signalled risk and restricted social activities had unintended consequences for students' education, even when the policies were designed to try to mitigate the impact on schools.

In the next section we summarise key literature in these fields. Section 3 outlines the pandemic and policy environment during our study period. Section 4 describes our data and estimation strategy. Section 5 describes our results in detail and Section 6 provides discussion and conclusions.

2. Existing evidence

There is an extensive existing literature in educational research on the role student background and other factors play in affecting absenteeism. Most of this research is descriptive in nature, either providing qualitative evidence or partial correlations between absenteeism and various student characteristic. Surveys and meta-analyses of this evidence are available, see for example Sosu et al (2021) and Gubbels, van der Put and Assink (2019). There are few strong conclusions from this evidence other than that absenteeism correlates with low income and otherwise disadvantaged backgrounds, and associated factors like drug use, family problems and low prior achievement. The reasons why these relationships emerge remain elusive. Policy solutions have included a range of measures, both punitive - fines and threats of prosecution – and motivational – like parent engagement and school community groups. There is a need for more evidence about which policies work. [1]

What are the consequences of school absences? There is good evidence about the negative correlation between absences and educational attainment, but much less causal evidence, though it has improved in recent years (Baker et al. 2022). It is difficult to estimate a causal effect because low attendance might reflect a range of factors that also influence low educational attainment.

There are a number of approaches that shed light on this question. Several studies use within student, between-grade (or subject) variation in absence and performance, with recent examples including Cattan et al. (2021) for primary schools in Sweden and Liu et al. (2021) for secondary schools in California. Despite the very different contexts and time periods, they both come to similar conclusions about the effect of school absence, finding that 10 days of absence reduces academic performance by 3 to 4.5 per cent of a standard deviation. Gottfried (2010) uses home-school distance as an instrument for attendance, finding that a one standard deviation increase in attendance (17 days) leads to a 28 percent of one standard deviation increase in GPA, though home-school distance is correlated with many student background

[1] Dee (2023) cites evidence on effective school-wide strategies which include providing engaging, culturally relevant instruction and school-based supports such as free meals, health care (e.g., asthma management), and social services. He also suggests that a promising school-wide practice is to engage and inform families about their child's school attendance.

characteristics so arguably not an ideal instrument for mitigating biases from unobserved student confounders.

Other studies use shocks to school attendance to identify effects on achievement. These include closure due to snow days (e.g. Goodman, 2014), flu outbreaks (e.g. Aucejo and Romano, 2016), teacher strikes (e.g. Baker, 2013, Belot and Webbink, 2010, Jaume and Willén, 2019, Johnson 2009, 2011) and riots (Montebruno, 2020). There are also studies that investigate the effect of lost instructional time due to adjustments in the length of the school year (Pischke, 2007), or variation in teaching hours within the week (Lavy, 2015).

For the most part, these studies do suggest a negative causal relationship between loss of instructional time (for whatever reason) on student educational outcomes. But the magnitude of the effect varies across contexts and can either be very large or be modest (even negligible). For example, relatively large effects are found from teacher strikes lasting about 6 weeks in Belgium (Belot and Webbink, 2010), of about 20 per cent of a standard deviation. In the context of student riots in Chile, Montebruno finds that 10 days of lost schooling costs students around 13 percent of a standard deviation in achievement (Montebruno, 2020). On the other hand, some studies find no effects of (fairly short) school closures due to snow (Goodman, 2014) or no effects on labour market outcomes for shortening the school year in Germany (Pischke, 2007). A recent study for Norway (Baker et al. 2022), evaluating the impact of a law designed to incentivise attendance in high school, found that large effects on attendance only had a mixed impact on measures of subsequent educational achievement.

It seems likely that the effect of school attendance depends on the duration and the counterfactual (i.e. what countervailing measures are put in place), as well as what groups of students are affected by any shock or policy change. Much work has considered what happened to students during the COVID-19 pandemic in the UK as well as internationally (e.g. see Farquharson et al. (2022) for a summary in the UK context). As the first period of school closures (from March 2020) took everyone by surprise, there was little guidance and resources for schools about online delivery. There was huge inequality in the extent of school engagement. For example, among the richest fifth of parents, nearly 60% of those sending their children to state schools reported that their child was providing online classes, falling to

40% among the poorest fifth of parents (Andrew et al. 2020). There were also huge differences among those sending their children to state schools and private schools, with the latter being much better resourced (Elliot Major et al. 2020). During the second period of closure (January-March 20201), schools' provision of learning evened out to some extent (Cattan et al. 2021). Throughout the pandemic, there were also huge socio-economic differences in access to reliable internet and home computers (Sutton Trust, 2021) and in home environment and resources, for example in the extent to which parents had the time and resources to engage with their children's learning.

There is a growing body of evidence on the effects of COVID-19 on educational achievement in the UK and internationally. For example, see Patrinos et al. (2022) for a recent survey of evidence, showing very large effects of the pandemic on global learning loss, though there is large variation between countries and between socio-economic groups within country. Another review of international evidence by Elliot Major et al. (2024) suggests that children suffered up to 6 months of learning loss during the pandemic, with children from low-income backgrounds experiencing an extra 2 months of learning loss. In the UK, most studies found that the first period of school closures in England cost children 1-2 months of expected progress, with larger impacts in maths (Rose et al. 2021; Renaissance Learning and EPI, 2021). Estimates of the effect of restrictions in 2021 suggest this cost primary school pupils around one month of expected progress (Renaissance Learning and EPI, 2021). Milanovic et al. (2023) find that there are some enduring effects of the pandemic on primary school attainment, particularly for younger students and in some subject areas (literacy – grammar, punctuation and spelling or GPS). Also, the gap between those classified as disadvantaged (eligible for the pupil premium) and other students has increased each autumn between 2019 and 2022 for primary school (Year 6) English and maths.

Our work is most closely related to that of Dee (2024), Jack et al (2023), Goldhaber et al (2023) and Ross et al (2024) for the US, who look at the relationship between the incidence of remote schooling during 2020 and subsequent student outcomes. Dee (2024) shows that the recent state-level growth in chronic absenteeism in the US is correlated with the share of schools that closed for in person teaching during 2020-21, and with levels of absence pre-pandemic. Jack et al (2023) and Goldhaber et al (2023) show that achievement and progress was lower on average in states that switched from in person to remote learning; the second paper uses pupil level data to show that students in high poverty schools fared worse as a result. Ross et al (2024) look at the effects student enrolment for in-person, rather than remote teaching on test scores in Connecticut schools in 2020-21, when attendance in person was very low. They find that students with more days in-person perform better on within-year tests, and that students in classes where a higher share of students attend in-person perform better, a feature they attribute to the challenges of teaching classes where there was a hybrid of in-person and remote learning. Our work differs from all these papers, in that we do not look at the effect of schools providing remote rather than in-person teaching, because this was not a systematic policy in England. We look instead at the effect of local area policies that influenced school attendance rates, even when schools were open to everyone for in-person teaching.

3. Schools and COVID-19 policy background

The analysis that follows investigates the influence of COVID-19 policy on pupil absences during the autumn term of 2020, and the subsequent effect on absences and achievement in the next academic year 2021/22. To see this period in the context, Figure 1 presents a weekly calendar of the sequence of events in England between the start of the pandemic in March 2020 and the period when the country was emerging from restrictions in spring 2021. The calendar shows the events happening in schools (in blue) and in the wider policy environment (in orange) in the weeks and months over this time period. The calendar shows the period up to the time when most restrictions on school opening had been lifted, though this was not the end of all restrictions generally. Restrictions were not fully lifted until July 2021. Even then, guidance on mask wearing persisted and people were still required to self-isolate if they tested positive for COVID-19. There was a renewed period of anxiety in December 2021 due to the Omicron variant, which mandated mask wearing and a COVID pass proving vaccination for entry to some venues. In February 2022, the requirement to self-isolate ended and free testing for most people stopped in April 2022, effectively signalling the end of the pandemic.

An important feature to note, is that schools were completely closed for around 17 weeks in this period, to all children other than those of parents designated as 'key workers'. The definition of key workers was wide-ranging definition, but covered health care workers, teachers and others involved in running of crucial services. These closure periods occurred in the first and third national lockdowns. At other times, schools were open to all pupils, although there were various restrictions at different times on the ages that could attend and on the operational processes in teaching schools. As discussed above, there was very little formalised remote learning during 2020, although practices varied widely.

We consider policy events which had the potential to generate variation in absence rates across schools during autumn 2020, primarily the assignment of Local Authorities (LAs) to different 'Tiers' which restricted social and economic activities in the area. We also consider the different guidance provided by LAs to schools on whether or not to follow central government advice to re-open to most students from June for the last few weeks of the 2019/20 academic year. Although this guidance was for the summer term of 2019/20, it may have influenced family attitudes to attendance on the return to school in the 2020/21 academic year, so we include it in our set of predictors of absence in autumn 2020.

The autumn term of 2020 coincided with a system of 'Tiers' that was introduced with different levels of restrictions in LAs, according to government assessment of the risks of Coronavirus transmission. There were initially three tiers: Tier 1, Moderate Risk; Tier 2, High Risk; and Tier 3, Very High Risk. A fourth, Tier 4, Stay at Home, was introduced in mid-December 2020 after the school term had ended (see Figure 1and Figure 2). The decision on when and where to introduce different tiers was based primarily on levels and rate of change of local COVID-19 infections. The early autumn term of 2020 also featured local lockdowns, which affected only a few areas, but imposed restrictions similar to Tier 4 and the national lockdowns. In all tiers, those who could work from home were advised to do so and people were told not to meet in groups of more than six, indoors or out (unless within the same household). As the level of the tier increased from Moderate (1) to Stay at Home (4), restrictions on social activity increased, banning meeting indoors, then outdoors and then meeting people outside of the household (or a recognised 'support bubble') anywhere. Leisure and hospitality services faced increasing restrictions and finally closure as the tier rating increases. In the Very High Risk and Stay at Home Tiers, non-essential travel outside the tier zone was forbidden. In summary, the main purpose of the tiers was to signal the level of risk and to restrict gatherings for social and leisure reasons. Although the tier designation did not affect schools directly – they were always open – there are good reasons to think they may have influenced whether or not families decided to send their children to school. Firstly, there was clear signal of the risk of infection at a time when there was great fear about the consequences of COVID-19. Secondly, as hospitality and leisure closed, there would have been more families working from home – especially those in low-income hospitality occupations - which could have encouraged children to stay at home too. We show in a companion paper, Gibbons, McNally and Montebruno (2024a), and later in this paper that absence rates were higher in schools in higher level tiers compared to Tier 1 and that the impact of the Tiers was greater for those from more income-disadvantaged family and neighbourhood backgrounds. It is this influence from the Tier designations that we use to generate systematic variation in absenteeism in the empirical analysis presented below.

4. Methods and Data

4.1 Data

Our main data source for the analysis is the Department for Education's National Pupil Database (NPD). This data contains an annual census of state school students in England, linked to their test results in assessments that form part of the National Curriculum and information on each pupil's absence over the term. The data set spans all years back to 2002, but for this work we use only two years either side of the start of the pandemic: 2017/18, 2018/19, 2020/21 and 2021/22 (the latest year available at the time when the data was made available to us). There is no data at all available for most of 2019/20, and no test results available for 2020/21 due to the constraints during the early stages of the pandemic and because information was censored to avoid disclosure relating to school activities.

Part of the data set is an annual census which records information on pupils' demographic characteristics, such as gender, age, ethnicity, free school meal entitlement, first language and an index of deprivation related to their home neighbourhood – the Income Deprivation Affecting Children Index (IDACI). The IDACI score is the proportion of under-16s living in income deprived households. The data on academic achievement records information on tests carried out at age 6-7 in when a student is in year group 1 (year groups are like US 'grades'), at age 11-12 in year group 6 at the end of primary school, and at age 15-16 in year group 11, when a student ends compulsory secondary schooling and takes exams leading to final qualifications, mainly General Certificates of Secondary Education (GCSEs). Students can take a range of different subjects at GCSE level. These periods in a child's school career are referred to as Key Stage 1, Key Stage 2 and Key Stage 4 in the National Curriculum structure, and we use this terminology when discussing test scores. For Key Stage 1, the information is limited to categories of achievement. We use this information only as a control variable. At Key Stage 2, we have test scores in three curriculum subjects: mathematics; grammar punctuation and spelling; and reading. At Key Stage 4, we use a total points score, which is based on adding up the points a student scores on each of their subjects. In all cases, we standardise test scores to percentiles in the student distribution in each year.

Individual pupil absence data also comes from the NPD. Pupil absence is recorded as the number of 'sessions' a pupil has missed in each term of the year. A session is half a school-day. The data does not show on which days these sessions were missed, or when they were missed within the term. There are various categories of absence shown in the data. There is an overall figure, which in pre-pandemic years refers to absences for any reason. For 2020/21 and 2021/22 there is an additional category defined as 'code-X' absences. This category refers to cases where absence was caused by pupils being unable to attend school because a household member or social contact had tested positive for COVID-19, which meant the pupil was barred from going to school and is most relevant in 2020/21. This kind of forced absence was self-evidently zero before the pandemic, declined sharply during 2021 and stopped being relevant early in 2022 when isolation was no longer a requirement. We refer to these as 'forced' absences and to other absences as 'unforced' or voluntary (although this is not meant to imply that there were not valid personal reasons for absence). Absence rates for each pupil are constructed by dividing the number of sessions absent by the number of sessions available at the school in the corresponding term, which is provided in the NPD data. Forced absence rates during the COVID-19 period are defined by the number of sessions missed due to self-isolation rules divided by the total sessions available. Unforced absence rates are defined by the overall absence rate (which excludes the forced absences during the COVID-19 period), by the sessions that are available *after* deducting those enforced for reasons of self-isolation. Absence rates are expressed as percentages throughout.

Data on the policies in place in the school's local area during 2020 are derived from various web sources. Guidance from LAs to schools on re-opening during summer 2020 was derived from LA web sites. We place LAs into two categories, those that advised schools to follow government policy and re-open during June 2020, and those that did not. This latter category includes those that advised schools not to open, advised schools to make their own decision, or else provided no guidance. Information on when LAs were placed in different Tiers during the period of Tier Regulations in autumn 2020 is derived from web sources, including government documents, legislation in Statutory Instruments, and Wikipedia. Given the annual structure of the NPD data, we construct variables for the intensity of exposure to the different Tiers or Local Lockdowns during autumn 2020 from the number of days an LA spent in each category between the start and end of the autumn 2020 term (and the number of days spent in Tier 4 after the end of term, for an additional test described later). Pupils are assigned to all these policy categories based on the LA in which their school is located.

Other data merged into the NPD data includes local COVID-19 infection rates, COVID-19 death rates and unemployment claimant rates (as a measure of the impacts of the pandemic on economic activity). These variables are all linked to a school at Middle Layer Super Output Area (MSOA) level, based on school location. MSOAs are census units containing an average of 7700 people.

4.2 Estimating the influence of pandemic absence on subsequent absence

Our regression specification for estimating persistence of absence from one year to the next takes the form:

$$autumn\_abs_{ist+1} = \beta_1 autumn\_abs_{ist} + \beta_2' controls_{ist} + f_s + f_i + v_{ist} \quad (1)$$

This specification is estimated on pupil level data, for 7 cohorts (year groups), spanning two periods *t*, pre and post pandemic. The variable $autumn\_abs_{ist}$ is pupil *i*'s absence in autumn term of year *t* (either 2017 or 2020 in our data) and $autumn\_abs_{ist+1}$ is pupil i's absence in the following autumn term (of 2018 or 2021). Absence is measured as the proportion of available sessions (half-days) missed by the pupil in the term (for reasons other than COVID infection). The set of control variables $controls_{ist}$ contains a range of pupil, school and local area characteristics: indicators of gender, ethnicity, free school meal entitlement, English first language, month of birth, birth cohort (i.e. school year group) and variables representing home deprivation (IDACI), COVID infection and death rates, and unemployed claimant count (at MSOA level). Factors f_s and f_i represent school and pupil level unobserved confounders, which we eliminate as school and pupil fixed effects in our regressions. The error term v_{ist} represents all other unobserved factors affecting a pupil's absence rate.

Given this setup and the years used in the sample (t = {2017, 2020}, t+1 = {2018, 2021}), equation (1) is equivalent to regressing the pupil-specific change in absence between autumn 2018 and autumn 2021, on the change in their absence between autumn 2017 and autumn 2020. The main parameter of interest, β_1, is therefore estimated as the average effect of the change in a pupil's absence between the pre-pandemic autumn 2017 term and autumn 2020 – caused, in part, by the pandemic policies in place in 2020 – on the change in absence between pre-pandemic autumn 2018 and autumn 2021, when most pandemic-related restrictions had been removed. It is from this parameter that we will infer the persistence of absence during the pandemic to absence post-pandemic.

A well-known problem with regression specifications like equation (1) estimated on data with a short time series component, where the dependent variable and lagged dependent variable are generated by the same stationary random process, is that estimates of β_1 will be downward biased by mean reversion (i.e. an upward shock to absence at time t will be followed, on average, by a downward shock to absence at time t+1). This issue has given rise to a large econometric literature on methods for estimating this kind of dynamic panel data model (Bun and Sarafidis 2015).

In our setting, we are interested specifically in the changes in pupil attitudes to attendance and the consequent changes in absence rates caused by COVID-19 policies, which would shift absence rates in ways that were not inherently mean reverting. In addition, there may be unobserved time-varying confounders, which lead past and future changes in absence to be correlated for other reasons than changes in pupil behaviour induced by the COVID-19 policies. We, therefore, mitigate the biases from both mean reversion and unobserved confounders by estimating equation (1) using a using a Two Stage Least Squares/Instrumental Variables (2SLS/IV) procedure, in which we predict autumn 2020 absenteeism in equation (1) from the autumn 2020 Tier Regulations, which serve as instruments. The assumption behind this approach is that the autumn 2020 Tier Regulations only affected future absence rates through their impact on pupil and family attitudes to attendance and contemporaneous absence rates.

In practice, rather than using the average effects of the policies across all pupils as instruments to predict autumn 2020 absenteeism, we interact the policy variables with a pupil background characteristic x_i. In this way, we identify the effect of absence on future absence from the variation in response to the COVID-19 policies across pupils who differ in terms of x_i. In our main results, we use the Income Deprivation Affecting Children Index (IDACI) of the pupil's place of residence for this interaction, because we expect the reaction to the autumn 2020 Tier Regulations to have differential effects according to the socioeconomic status of a pupil and their neighbourhood, as shown in our previous work (Gibbons, McNally and Montebruno 2024). In fact, poverty has been shown to directly relate to student absences in previous research (Gennetian et al. 2018; Gottfried et al., 2014). The advantages of this approach are that: it generates more variation in predicted absenteeism (i.e. our instruments are stronger); it means that we can control separately for the main, average effects of the COVID-19 policies on absenteeism in equation (1). Our first stage regression is thus of the form:

$$autumn\_abs_{ist} = \delta_1{}'(policy_{st} \cdot x_i) + \delta_2{}'controls_{ist} + f_s + f_i + \varepsilon_{ist}$$

in which $policy_{st}$ is a vector of variables for the number of days a school spent in each of the tiers under the autumn 2020 Tier Regulations, plus an indicator of the Local Authority guidance on re-opening in summer 2020, as described in section 4.1. When using this 2SLS method, the vector of policy variables $policy_{st}$ is also included in the set of control variables $controls_{ist}$ in equation (1) and (2).

4.3 Estimating effect of pandemic absence on subsequent test scores

The method for estimating how pandemic absence affects future key stage achievement follows a similar strategy to that set out for future absence above. The main regression specification takes the form:

$$ks_{ist+1} = \gamma_1 autumn\_abs_{ist} + \gamma_2' controls_{ist} + f_s + v_{ist} \quad (3)$$

where ks_{ist+1} refers to student i's Key Stage 2 or Key Stage 4 tests (converted into percentiles in the student distribution in each year). Likewise, this regression is estimated for two periods t, corresponding to the pre and post pandemic periods, t = {2018, 2020}, t+1 = {2019, 2021}. In this case, we have only two cohorts of students for each key stage level, those taking the Key Stage 2 tests in 2019 and 2021, and those taking the Key Stage 4 (GCSE) tests in 2019 and 2021. Referring to Figure 3, cohort 5 forms a treated cohort (experiencing the autumn 2020 pandemic regulations) and cohort 8 the pre-pandemic control cohort for Key Stage 2. Cohort 10 and cohort 13 are the corresponding cohorts for Key Stage 4. The coefficient of interest here is γ_1, the effect of absence on future achievement. Given our interest in the effect of pandemic policy induced absence on future outcomes, we use a similar first stage regression to equation (2) to predict shocks to autumn 2020 absence induced by autumn 2020 pandemic regulations and estimate (3) using a two stage least square procedure.

Note, since we only observe student outcomes in specific key stage level once (they only take the tests once during their schooling), it is not feasible to control for student fixed effects when we estimate equation (3). Since our instruments in the first stage regression are interactions between geographically targeted policies and student's home deprivation (IDACI) score, there is some risk that estimates of γ_1 could be biased by any direct influence of deprivation on key stage scores, if this effect varies by geographical location. To address this concern, we extend the specification in (3) to control for the interactions between IDACI scores and school fixed effects $f_s x_i$ (i.e. a different pupil IDACI score control variable for each school in the data).

5. Results

5.1 Descriptive statistics

The crucial underlying feature of our study is the change in pupil absenteeism over the pandemic period. These levels of absenteeism for the two periods in our data – pre-pandemic (2017/18-2018/19) and pandemic (2020/21-2021/22) are shown in Table 1. The figures show the percentage of 'sessions' missed in the autumn term of each year, a session being a half-day. Given there are around 100 possible sessions in a typical term, the numbers can also be interpreted as the number of sessions missed per term. The first three rows show levels of absence in the second year of each period (2018/19 or 2021/22). The second three rows refer to the levels of absence in the first year of each period, i.e. the time lagged absence (2017/18 or 2020/21). In each group of three rows we report means and standard deviations for: (1) overall absence level; (2) the absence that was forced because of the need to self-isolate after contact with people testing positive for COVID-19 ('code X' absence in our data); and (3) unforced absence that was not directly due to self-isolation or illness from COVID-19, but could be due to other illness. We exclude absence directly due to COVID-19 illness, although these numbers are actually very small given the children were relatively unaffected by the virus.

These figures show that overall absence was 3.5-3.7 percent in the autumn terms prior to the pandemic, all of which was unforced according to our definition. In the autumn term of 2020, at the height of the pandemic and the policies aimed at reducing interpersonal contact, overall absence rose to 13.5%, 5.6 percent being unforced absence and 8.7 percent due to the requirement to self-isolate. In the autumn term of 2021, overall absence had fallen to 9.3 percent, but this was due to the drop in the numbers unable to attend school because they were self-isolating. Average unforced absence increased to 7.7 percent, more than double what it was pre-pandemic. Clearly there was a big change, and the change is due to increases in absence for reasons unrelated to COVID-19 infection. Later, when discussing our regression results, we present even more pronounced increases in the proportion of children missing significant amounts of the school term.

Other descriptive statistics for our estimation dataset are tabulated in Appendix Table 12. These figures are for the pandemic years of our data. The first two columns summarise our absence data set that covers all the cohorts in our data. The second two columns describe the KS2 dataset, i.e. the cohorts taking their KS2 tests in 2019 and 2022. We will not describe all these figures in detail. The most relevant relate to the policies that were in place in 2020 and the way these impacted on schools. Nearly 28% of schools were in LAs that did not advise their schools explicitly to follow government advice on re-opening in summer 2020. As it turns out in our analysis, this had little impact on subsequent absence behaviour, but we report here for information. The numbers relating to the time schools and their pupils spent under different levels of regulation in autumn 2020 are more relevant. Schools spent an average of 7.4 days under local lockdown, a figure that is driven by a small number of schools in a few LA being under lockdown for extended periods. During the first half of the term, the period of the 1st Tier Regulations, schools spent an average of 10.6 days in Tier 1 (least restricted), 9 days in Tier 2 and 2.4 days in Tier 3 (most restrictions). This was followed by two weeks of national lockdown, which we do not use in our analysis as it applied equally to all schools. During the second half of the term, schools spent on average, 0.2 of a day in Tier 1, 8.6 days in Tier 2 and 8.2 days in Tier 3, after which schools closed for the Christmas holidays.

Looking at attainment, these descriptive statistics are not very informative, because our variables are standardised to represent student percentiles in the distribution in each year. Published statistics are more informative about the general changes over this period. [2] Between 2018/19 and 2021/22 the percentage of students reaching the expected level in KS2 fell from 79% to 71% in maths, and from 78% to 72% in GPS (grammar, punctuation and spelling). They remained roughly the same between 2022 and 2023 (the most recent year for which figures are available, but beyond our study period). Prior to the pandemic, the proportion reaching these expected levels had been rising steadily year by year. Interestingly, performance in reading was largely unaffected by the pandemic, with 73-75% of students reaching the expected levels both before and after.

Key Stage 4 (GCSE) performance during the pandemic is difficult to measure on a comparable basis with previous years, because grades were adjusted in 2019/20, 2020/21 and 2021/22 to compensate for the challenges students faced over the pandemic, so as to not compromise future education and career paths. Official statistics show that the average 'attainment 8' score (the average of the score on a student's best performing subjects) increased from around 46.5 in the three years preceding the pandemic to over 50 points in 2019/20-2020/21 and 48.8 points in 2021/22. The adjustments to these scores mean we do not focus on KS4 outcomes in our main analysis, although we look at them briefly when we discuss the

[2] See https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment-national-headlines#releaseHeadlines-charts

regression results. Work on KS4 will require data from 2022/23 onwards, when there was a return to pre-pandemic grading standards.

5.2 Persistence of absence

5.2.1 Baseline results

Regression results relating to the effect of absence in autumn 2020 on absence in autumn 2021 are shown in Table 2. The table reports regression coefficients and standard errors corresponding to the coefficient $\beta_1$in equation (1), which is estimated on the pupils in birth cohorts 4-10 in year groups 1-11 as defined in Figure 3. Standard errors are adjusted to correct for heteroscedasticity at school level and correlation between pupils in the same school (i.e. they are clustered at school level). All the regressions control for school and birth-cohort fixed effects, i.e. all factors which are constant at school and birth cohort level over the years of our sample. All regressions also include a post-pandemic dummy variable, which controls for any changes between the pre-pandemic (2017-2019) and post pandemic (2020-2021) periods that are common across all schools.

The coefficient shows the magnitude of the effect of one session of absence in the autumn term of a given year on the number of sessions a pupil was absent in the autumn term one years later. A coefficient of zero would imply that a change in autumn absence is transitory and has no impact on absence next year. A coefficient of one would imply that any change in autumn absence is permanent and is still observed in the autumn next year. Intermediate values between zero and one represent increasing levels of persistence. Values above one would imply that changes in absence in one year are amplified and lead to bigger changes in subsequent years. Values below zero, would imply that an increase in absence in one year leads to a reduction in absence in subsequent years.

Table 2 is organised with ordinary least squares (OLS) regression results in columns 1-3, and two stage least squares (2SLS) results, where we predict autumn 2020 absence from the days each school spent in each Tier of the Tier Regulations, in columns 4-6. The first stage corresponding to this prediction is shown in Table 3, to which we refer later in the text. Columns 1 and 4 include school fixed effects, a post-pandemic dummy and pupil cohort fixed effects, but no other control variables. Columns 2 and 5 add in pupil characteristics, local unemployment rates and COVID-19 case and death rates, as set out in the table notes. Columns 3 and 6 control for pupil fixed effects. In the pupil fixed effects estimates, the coefficients are estimated from the pupil-specific changes in absence rates between the pre-pandemic and post-pandemic periods.

In column 1, the coefficient on last year's autumn absence is approximately 0.3, implying that missing 10 percent of the available sessions in an autumn term is associated with 3 percent of sessions missed in the next autumn term. Although we are only using autumn term data here, we assume this result could be generalised to other terms and absence over an entire year. There are, however, good reasons to think this relationship might not be causal. Unobserved pupil-level attributes (confounders) that affect absence would lead absence in one year to be correlated with absence in the next year, even if there was no causal relationship between changes in absence one year and changes the next. Conversely, as discussed in Section 4.2, mean reversion will tend to downward bias the coefficient, as if a pupil has randomly high pupil absence one year they will have, on average, randomly low absence the next.

Adding the set of control variables into the regression in column 2 makes little difference to the coefficient. In column 3, we control for pupil fixed effects (i.e. pupil specific constants) which account for all non-time varying pupil characteristics, further mitigating biases from unobserved confounders. The disadvantage of adding more control variables and pupil fixed effects is that it will tend to exacerbate downward bias from mean reversion and, indeed, in column 3 the coefficient on past absence drops to 0.183. These OLS results then suggest only a moderate role for higher absence in one year causing higher absence in the next, but we suspect these coefficients may be considerably downward biased by mean reversion.

To better address these potential biases, we turn to 2SLS estimates in columns 4-6. The 2SLS estimates in columns are, in effect, averaging autumn 2020 absence rates for pupils in groups corresponding their LA and their IDACI score, weighted by the number of days the LA was in each tier. This averaging, and the fact that the policies constitute an exogenous shock that is not part of the normal process governing pupil absence, will mitigate the mean reversion problem. Here we see the coefficient on autumn absence increases substantially towards 0.654 in the pupil fixed effects estimates in column 6.

The interpretation of this result is that a 10 percent increase in absence induced by the local 2020 pandemic policies persisted as a 6.5 percent increase in absence in 2021. Extrapolating this to future years outside those in our dataset implies that the impact of the 2020 policy shock would take 7 years to erode to 5% of its initial value (because $0.65^7 = 0.049$).

The table also includes some tests relating to the validity of the 2SLS method. The F-test of the excluded instruments refers to the relevance of our instruments (the policy-IDACI interactions) in predicting autumn 2020. The standard rule of thumb is that F-statistics above 10 are acceptable, and all the values in the table are well above this. The overidentification test refers to the Sargan/Hansen test of the overidentifying restrictions. The null hypothesis is that the separate instruments have the desirable property of implying the same coefficient in the second stage. In columns 6 and 7, these tests have marginally acceptable p-values of 0.07, so we would reject the null only at the 5% level. In column 8, with pupil fixed effects, the test is much more reassuring about the validity of our specification, with a p-value of 0.278 (Windmeijer, F. (2019)).

Table 3 reports the first stage corresponding to the 2SLS estimates in Table 2. These results tell us about how the pandemic policies of 2020 influenced pupil absence, so are interesting in their own right. Column 1 shows the coefficients from the baseline effects of the policies on the least deprived (IDACI=0) group. Column 2 shows the coefficients relating to the interaction between the policy and a pupil's home IDACI score, the excluded instruments in our 2SLS method.

Most of the coefficients in column 1 are small and statistically insignificant, implying that there was little baseline effect of the various Tier regulations and summer 2020 LA re-opening guidance on individual pupil absence for those from the least deprived neighbourhoods. The exception is the number of days spent in Tier 3, the highest tier of the 2nd Tier Regulations which were in place for the second half of the autumn 2020 term, though the effect is not large. Eight days spent in Tier 3 (which is the mean, from Table 12) implies a 1.1% increase in absence relative to the lowest Tier 1 category, or around half a day (given there are around 50 days in the term).

In column 2 we see large, strong effects from the interaction of the number of days spent in Tier 3 or Tier 2 and a pupil's home IDACI (deprivation) score. In other words, the number of days spent in these tiers during the second half of autumn term had very different effects depending on the neighbourhood the pupil lived in. Given the scale of the IDACI score – the proportion of children in low-income households - the coefficients of 0.44 imply that 8 days in Tier 3 or Tier 2 induced a pupil in a neighbourhood with 100% deprivation to have an absence rate 3.5 percentage points higher than a pupil in a neighbourhood with no deprivation (8 x 0.44). Another way of looking at this is that if 8 days in Tier 2 or Tier 3 induced half a day of absence over the term for the least deprived (the baseline), it will have induced over 2.25 days of absence for the most deprived. The response at the mean of IDACI (0.18) to 8 days in Tier 3 is a 1.8 percentage point increase in absence rates, or just under a day ((0.144+0.18*0.44)*8). For 9 days in Tier 2 (the mean), the increase in absence rates would be 1.3 percentage points ((0.062+0.18*0.44)*9). A pupil spending the average 9 days in Tier 2 and 8 days in Tier 3 in the second half of autumn term would have had 3.1 percentage points (one and a half-days) more absence over the term than those in the relatively unrestricted areas. These are big effects, given the baseline mean absence rate of 3.5 percent in the pre-pandemic period of our data. As we shall in later analysis, the implications for chronic absence are even more severe than these average marginal impacts seem to suggest. The interactions of IDACI with the other policy classifications are much smaller and/or insignificant, implying no detectable effect on absence rates in 2020. The 2nd Tier Regulations in the second half of autumn 2020, which followed the 2nd National Lockdown in November, seem to have a uniquely important effect on absence rates.

Columns 3 and 4 in Table 3 offer a test of the credibility of these first stage regressions in showing a causal relationship between the Tier Regulations and pupil absence. Here we include the number of days spent in Tier 4 and its interaction with IDACI, Tier 4 being a designation that only began mid-December 2020 after the school term had ended. Ideally, we should see no impact on autumn 2020 absence from time spent on Tier 4, given it post-dated the term. While the coefficient is moderate in magnitude it is much smaller, only marginally statistically significant and of opposite in sign to the effects of Tier 2 and Tier 3 suggesting this is likely just to be a spurious result, or due to some reverse link between behaviours in the autumn term and whether or not an area subsequently went into Tier 4 (e.g., if family behaviour in staying home and keeping kids off school reduced the possibility of entering Tier 4). Either way, the result suggests that, as we would expect, there is no causal link from Tier 4 days to autumn term absence.

Although tangential to our main research question, the regressions in Table 2 also shed light on the relationship between student background characteristics and absenteeism. In Appendix Table 14, we report the coefficients on the control variables relating to student characteristics in Table 2, column 2, to illustrate these relationships (which are conditional on autumn absence the previous year). The key features of these results are that absenteeism higher for low-income students (by 2.3 percentage points) and those from deprived neighbourhoods (3.6 percentage points higher for the most deprived compared to the least), higher for white-British students than other ethnicities (by around 2 percentage points compared to Asian and Chinese students), higher for those with English as a first language (by 0.5 percentage points) and higher in the pandemic year 2021/22 than before (by 3 percentage points). There is an interesting pattern for the month of birth dummies, with younger students marginally less likely to be absent than their older peers throughout a year. Girls have slightly higher absence rates than boys. COVID case rates in 2021 were, unsurprisingly linked to higher absence in 2021. The coefficient on case rates in 2020 is negative, but this coefficient is hard to interpret given case rates at this time are a strong predictor of absence rates in autumn 2020, which are already controlled for in the regression. In the first stage regressions of autumn absence on these controls (unreported in the tables), the coefficient on the autumn 2020 case rate is 1.266 with a standard error of 0.109, indicating that an increase of 1 case in 100,000 in autumn 2020 was associated with a 1.2 percentage point increase in autumn 2020 absence.

5.2.2 Differences across student groups

So far, we have looked at the persistence of absence on average across the whole student population. In Table 4, we break this down by different student groups, along dimensions related to income, geography, sex and phase of schooling. The column headings show which group the estimate relates to. In these regressions we do not control for pupil fixed effects, so the comparable figure for the population is in Table 3, column 5. The figures show that in many cases, there is little difference across groups – high/low neighbourhood income deprivation, north versus south of England, male versus female. There are bigger differences between pupils with low-income families (FSM) versus other (not FSM), absence being twice as persistent from year-to-year for low-income pupils. There are some differences between large metropolitan areas (conurbations), smaller urban areas and rural areas, but these differences are not large. The biggest gap is between primary and secondary school pupils, with much a much bigger impact for secondary school pupils. The coefficient for primary school pupils implies that the impact of the pandemic policies on absence would have diminished very quickly over the years, falling to less than 2% of the initial change within 3 years. Based on these results we would not expect absence during the pandemic to have had long lasting effects on primary school pupils. By contrast, the effects are very persistent throughout secondary school.

5.2.3 Chronic absence

Much of the focus of academic and policy discussion on post-pandemic absence has been the rise in chronic absenteeism, meaning very high levels of absence for some pupils. We turn to this issue in Table 5. The basic specification is the same as Table 2, column 6, but we replace the dependent variable with a binary indicator, that is set to one if pupil's absence rate exceeds a certain threshold – either 10%, 20% or 40% in columns 1, 2 and 3 respectively. The first row of the table reports the coefficient on last year's absence rates, which should be interpreted as the effect of a one percentage point change in autumn absence on the probability of being absent for more than 10%, 20% or 40% the following autumn term.

The table also shows the means of the dependent variable, i.e. the proportions absent for more than 10%, 20% or 40% of sessions, in the 2019 (pre-pandemic) and 2021 (post-pandemic) periods. These figures are striking, and show that, indeed chronic absence rose sharply. The proportion absent for more than 10% of sessions over the term (i.e. one day every two weeks, or a week every term) more than quadrupled. The proportion missing 20% of sessions is seven times higher than before the pandemic. The proportion missing 40% (two days a week, or 4 weeks perm) is ten times higher.

The coefficients in the first row suggest that absence during autumn 2020 was a significant factor here. Recall, from Section 5.2.1, that the baseline increase in absence rates in autumn 2020 in response to the average number of days in Tier 2 or Tier 3 of the 2nd Tier Regulations was 2.9 percentage points (compared to the less regulated Tier 1 areas). This average change in autumn 2020 absence implies a 4.6-percentage point increase (2.9 x 0.016) in the proportion of pupils absent for more than 10% of the following autumn term, a 7.5-percentage point increase (2.9 x 0.026) in the proportion absent more than 20% of the time, and a 2.3-percentage point increase (2.9 x 0.008) in the proportion of pupils absent 40% of the time. In short, the knock-on effects of absence in autumn 2020 explain all (or nearly all) of the increase in over-20% and over-40% absence rates, and a substantial proportion of the increase in over-10% absence rates (the mean changes were 22, 9 and 2 percentage points respectively, from the second and third rows of Table 5).

5.2.4 Forced versus unforced absence

A crucial issue to remember is that schools were open during the autumn 2020 period, and much of the absence observed this term was not enforced by schools or by government. As we saw in the descriptive statistics, the absence rates categorised as compulsory – because pupils were self-isolating, were 8.7% and absence rates for other reasons which were not directly COVID related were 5.6%. A salient question then is to what extent the effects of autumn 2020 absence were due to compulsory absence or due to more 'voluntary' types of absence. We answer this question in Table 6. The regression specifications here are like those of Table 2, column 6, but we split autumn 2020 absence into an unforced component, and a component due to enforced absences due to self-isolation (code X absences). By construction, the self-isolation related compulsory absences were zero in the pre-pandemic period.

One technical challenge we have here is that we now have two potentially endogenous absence variables – unforced and compulsory absence. We tackle this by treating the compulsory absences as exogenous and unrelated to unobserved pupil, school and area characteristics, which is justifiable because the compulsory absences were determined by a fixed set of rules related to whether a pupil had been in contact with someone with COVID symptoms. We treat unforced absence as endogenous and instrument with the policy IDACI interactions as we did before. Column 1 shows ordinary least squares estimates, and the coefficient is broadly similar to what we saw in Table 2 for total absences. In column 2 unforced absence in autumn term again explains most of the effect we observed for total absence rates in Table 2, column 6. The effect of compulsory absence is negative, so, on the face of it, compulsory absence in autumn 2020 either encouraged pupil attendance in future terms – possibly to compensate for lost time - or made little difference. The take-away from this part of the analysis is that it was likely a shift in family attitudes to attendance during the autumn term of 2020, induced by the local public health policies, social and work restrictions of the time, that has persisted post-pandemic. Being prohibited from attending school seems to have had little lasting impact.

5.3 Impact on achievement

5.3.1 Baseline results for Key Stage 2

We now turn to the analysis of the effects of absence and lost schooling time on subsequent achievement. Table 7 reports the first baseline results for primary school pupils taking Key Stage 2 tests. The table reports regression coefficients and standard errors corresponding to equation 3, which is estimated on a dataset of two cohorts of pupils taking their KS2 tests in summer 2019 and 2022, the age-cohorts labelled 5 and 8 in Figure 3. There are separate regressions for each of the available tests: maths; grammar, punctuation and spelling (GPS); and reading. The test score outcomes are re-scaled to percentiles in the pupil distribution, so 100 corresponds to the top-ranking 1% of pupils and 1 the bottom ranking 1%. We report three sets of specifications for each subject. Columns 1-3 are ordinary least squares estimates, with school fixed effects (i.e. school specific constant terms) and a full set of control variables detailed in the table notes. Columns 4-6 are 2SLS estimates where we predict autumn 2020 absence from the COVID policy-IDACI interactions, again with school fixed effects and control variables. Columns 7-9 replace school fixed effects with school x IDACI score interactions, allowing the effect of IDACI on pupil test scores to differ for each school. In the specifications with school fixed effects, we are comparing pupils attending the same school. When we control for school-IDACI interactions, we are comparing pupils in the same school who have the same home IDACI score. An important point to emphasise, is that when we predict autumn 2020 absence from the COVID policy-IDACI interactions, all pupils in the same school with the same IDACI score get the same level of predicted absence in autumn 2020. Therefore, in the 2SLS estimates we are comparing a pupil with a given IDACI score, in a given school in the pandemic cohort, with another pupil with the same IDACI score, in the same school but in the pre-pandemic cohort. For reference, the first stages of these 2SLS regressions (i.e. the regressions predicting autumn 2020 absence) are shown in Table 8 in the Appendix and are similar to those described before for Table 3. The mean impact on autumn absence from being 9 days in Tier 2 and 8 days in Tier 3 of the 2nd regulations (the average duration) from the first stage regression of columns 1 and 2 is 1.45 percentage points ((0.048+0.392*0.18)*8+(-0.004+0.329*0.18)*9). One further point to note is that our check of whether the Tier 4 designation that post-dated the autumn term show any unexpected association with autumn absence is much more convincing: we find no significant effect.

Looking across the columns of Table 7, all the coefficients are negative, so it is immediately clear that lost time in school is associated with lower educational achievement, which should come as no surprise. The magnitude of these coefficients shows the average effect of a percentage point increase in autumn absence in school year-group 5 on a pupil's achievement in KS2 tests in year-group 6 (recall, we do not have 2021 test scores, so are unable to look at the impact of 2020/2021 absence within the same year as the tests are taken). Mean reversion in pupil absence rates again likely attenuates the ordinary least squares estimates in columns 1-3, which show large and significant effects from absence on future achievement. The coefficients indicate that 1 percentage point of absence (roughly half a day over a term) reduces a pupil's maths scores by 0.18 percentiles, GPS scores by 0.12 percentiles and reading scores by 0.06 percentiles.

The implied effects from the effect of the COVID policies on absence in columns 4-9 are much larger. The coefficients for maths and GPS scores are around one in columns 4-6 and increase to 1.1-1.3 when we better control for pupil and school characteristics in columns 7-9. The coefficient for reading is 0.3. These numbers mean that a 1-percentage point increase in absence reduces achievement by over 1 percentile in maths and GPS and 0.3 percentiles in reading the following year. As we saw above, the average increase in autumn 2020 absence associated with 9 days in Tier 2 and 8 days in Tier 3 was 1.45 percentage points. So, these Tier regulations came at an average cost of around 1.45 percentiles of achievement in maths and GPS, and 0.4 percentiles in reading. The smaller impact on reading might be due to pupils reading more when not at school, but we have no data here to verify that hypothesis.

The discussion above assumes that the pandemic policies affected pupil absence and this in turn affected subsequent achievement. A reasonable question about these findings is whether they really represent the effects of absence, or whether there are other direct effects from the pandemic policies which affected subsequent achievement. Examples of these alternative channels are the potential impact of the policies on children's mental health, or family financial resources, which could influence future achievement without affecting attendance at school. The assumption that the policies affected achievement via absence alone is untestable (without data on other child and family outcomes), but either way the results imply that the pandemic policies affected future achievement. One way to think about the effects of the policies on achievement is to multiply the coefficients in Table 7, by the first stage coefficients in Table 8. For example, the implied effect of 1 day in Tier 3 of the 2nd Tier regulations on maths scores for the least deprived children would be -1.120 x 0.048 = -0.054, whereas the implied effect on the most deprived would be -1.120 x (0.048 + 0.392) = -0.492 percentiles. A more direct way to estimate this is to regress the test scores directly on the variables representing the pandemic policies (the 'reduced form' regression). These regressions are shown in Appendix Table 13. Here, the effect of a day in Tier 3 on maths for the least deprived is near zero (0.005) and non-significant, whereas the effect for the most deprived is around -0.275 percentiles. Clearly there is some discrepancy between these two approaches, although the general conclusions are similar: the pandemic policies – Tier 2 and 3 in the second part of autumn 2020 in particular - reduced achievement for pupils from the most deprived neighbourhoods, but had little effect on the least deprived.

5.3.2 Differences across student groups

In Table 9, we split the sample into different student groups and re-estimate our regressions, to see if absence has a bigger effect on achievement for some groups than others. The table structure organises the groups into columns and the different subject test scores into different rows, but each number represents a separate regression. Looking at maths and GPS scores, the most striking difference is between students from disadvantaged backgrounds and other students, measured either by their home IDACI score or whether or not they are entitled to free school meals. Surprisingly, the effects are larger for the least deprived – i.e. a day of absence reduces test scores of the less disadvantaged students by more than those of the more disadvantaged students – though these differences are not always statistically significant. This is against a background of higher levels of absence, greater sensitivity of absence to the pandemic policies, lower levels and lower variability in attainment amongst students from more disadvantaged students. It is possible that the lower coefficients for the more disadvantaged students indicate that absence has a smaller impact on scores at the lower end of the KS2 score distribution, than it does at the top. The differences between other groups are less marked.

The picture for reading scores is quite different. We find no effect of absence on reading for students from the less deprived neighbourhoods, but a strong effect for those from the most deprived, though no difference between FSM-eligible children and others. Boys' reading scores are affected by absence, but girls are not. These differences are plausibly due to different reading habits outside of school. It is well-documented, for instance, that boys read less in their leisure time than girls do (e.g. See OECD, 2015).

5.3.3 Forced versus unforced absence

In Section 5.2.4 we looked at how forced and unforced absence in autumn 2020 affected subsequent absence and found the unforced absence – i.e. absence resulting from the decision of families to keep children away from school – has a much bigger impact than absence that was obligatory when students were self-isolating. In this section we look at how these types of absence differed in their effect on subsequent achievement. The regression results are in Table 10. The ordinary least squares in results in columns 1-3 show a similar picture to that for subsequent absence. Forced absence due to self-isolation has a much smaller effect than absence that is of a more voluntary nature: missing 1 percent of the term due to self-isolation reduced test scores at KS2 by 0.03-0.04 percentiles, whereas unforced absence is associated with a 0.1-0.21 percentile drop. In columns 4-6 we apply our 2SLS method where we predict autumn 2020 unforced absence from the pandemic policies. The effects of unforced absence are now very large. The numbers imply that an absence rate of 1% reduces reading scores by 1.8 percentiles, maths scores by 5.5 percentiles and GPS scores by 6.2 percentiles. These numbers seem implausibly large for the general effects of absence; but they may not be so implausible in this specific context, given we found that the average number of days in Tier 2 and 3 in later autumn 2020 increased absence in autumn 2020 by only 1.45 percent, implying that on average, Tier 2 and 3 reduced KS2 achievement by up to 9 percentiles (in GPS).

5.3.4 Impact of absence on Key Stage 4

As discussed earlier, Key Stage 4 outcomes are not measured in 2020-2022 on a comparable basis with pre-pandemic years, given that they were adjusted to compensate for the challenges students faced during the pandemic period. Consequently, they are unlikely likely to provide reliable indicator of the effects of pandemic on achievement. Nevertheless, for completeness, we show basic results for the effect of autumn 2020 absence on KS4 achievement in Table 11. The table has three columns, the first an ordinary least squares regression, with column 2 and 3 showing 2SLS results using our policy x IDACI interactions as instruments. All regressions include the full set of control variables. Column 1 and 2 control for school fixed effects. Column 3 controls for IDAI x school fixed effects. The dependent variable is the pupil's total KS4 total points score (converted to percentiles), and has a mean of 53.562 and standard deviation of 27.191.

The OLS estimates are negatively signed and statistically significant, but as discussed before, there are good reasons to expect this estimate to be biased by unobserved pupil confounders and mean reversion. The 2SLS estimates, where we predict absence from the policy variables, are all positive and statistically insignificant. The indication here is that there is no detectable causal effect of absence on KS4 achievement, though given that KS4 scores were adjusted to compensate for pandemic-related disadvantages, it is difficult to draw strong conclusions from this finding.

6. Discussion and conclusions

Variation in policy restrictions during the pandemic had large unintended consequences for student absence at the time and in subsequent years. This happened despite the fact schools were ostensibly open and we show that effects are not driven by students needing to self-isolate. This illustrates that the effect of the pandemic on subsequent absences and learning loss is not only attributable to compulsory school closures. Also, it does not appear that the disruptive effects of having to self-isolate (whenever another student became ill) had a lasting effect on future absences or achievement.

We find that local pandemic policies had heterogenous effects on absences across groups of students, being particularly large for disadvantaged students. This also feeds through into persistent absence (1 year later). This is one mechanism for the widening socio-economic gap in educational achievement during and after the pandemic. It also illustrates that more restrictive policies negatively affected those already facing hardships due to socio-economic deprivation. This shows that those needing most assistance to recover from the effects of COVID-19 are those from under-privileged backgrounds.

Even though absences are more prevalent in secondary school than in primary school, the effect of variation in policy restrictions comes out more strongly in subsequent primary school attainment (at age 11). This may be because of measurement problems in GCSEs or it may be because variation in policy restrictions really did have a more severe impact on younger pupils. While school attendance is very plausible as the mechanism through which policy restrictions impact on future test scores, there are other possible mechanisms (which will we explore in ongoing work). The strong effects observed for pupils in primary school suggests that the educational effects of COVID-19 will persist in the absence of effective policies to counter the effect of pandemic-induced learning loss.

7. References

Adams, R. (2024) Record numbers of pupils in England absent for long periods, DfE data shows, The Guardian, 21st March 2024: https://www.theguardian.com/education/2024/mar/21/record-numbers-of-pupil-absences-in-england-dfe-figures-show

Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., and Sevilla, A. (2020a), 'Family Time Use and Home Learning during the COVID-19 Lockdown', Institute for Fiscal Studies (IFS), Report 178.

Baker, J. Drange, N., Griefen, H.M., (2022). An Evaluation of a National Program to Reduce Student Absenteeism in High School. NBER Working Paper 20194.

Baker, M. (2013). Industrial Actions in Schools: Strikes and Student Achievement. Canadian Journal of Economics, 46(3), 1014–1036.

Belot, M., and Webbink, D. (2010). Do Teacher Strikes Harm Educational Attainment of Students? Labour, 24(4), 391–406.

Blainey, K., and Hannay, T. (2021), 'The Effects of Educational Disruption on Primary School Attainment in Summer 2021', RS Assessment from Hodder Education.

Bun, Maurice J. G., and Vasilis Sarafidis, 'Dynamic Panel Data Models', in Badi H. Baltagi (ed.), The Oxford Handbook of Panel Data (2015; online edn, Oxford Academic, 5 May 2015), https://doi.org/10.1093/oxfordhb/9780199940042.013.0003, accessed 11 July 2024.

Cattan, S., Farquharson, C., Krutikova, S., Phimister, A., Salisbury, A., and Sevilla, A. (2021b), 'Home Learning Experiences through the COVID-19 Pandemic', Institute for Fiscal Studies (IFS), Report 195.

Cattan, S., Kamhofer, D.A., Karlsson, M. and T Nisson (2021a). "The short- and the long-term effects of student absence: evidence from Sweden" IFS Working Paper 21/06

Dee, T.S. (2024) Higher chronic absenteeism threatens academic recovery from the COVID-19 Pandemic, PNAS, 121 (3) e2312249121

Elliot Major, L., Eyles, A., and Machin, S. (2020), 'Generation COVID: Emerging Work and Education Inequalities', Centre for Economic Performance (CEP), COVID-19 Analysis, Paper 011.

Elliot Major, L., Eyles, A., Lillywhite, E., and Machin, S. (2024). A generation at risk Rebalancing education in the post-pandemic era. Report for the Nuffield Foundation.

Farquharson, C., McNally, S., and Tahir, I. (2022). Education inequalities. IFS Deaton Review of Inequalities. Institute for Fiscal Studies. London.

Gennetian, L.A., C. Rodrigues, H. D. Hill and P.A.Morris (2018) Stability of income and school attendance among NYC students of low income families, Economics of Education Review 63: 20-30

Gibbons, S., S. McNally and P.Montebruno (2024) Pupil absence and the pandemic, working paper

Goldhaber, D., T. J. Kane, A. McEachin, E. Morton, T. Patterson and D. O. Staiger (2023), The Educational Consequences of Remote and Hybrid Instruction during the Pandemic, American Economic Review: Insights, 5 (3): 377–92

Goodman, J. (2014). "Flaking out: Student Absences and Snow Days as Disruptions of Instruction Time". NBER Working Paper, 20221.

Gottfried, M. A. (2010) Evaluating the Relationship Between Student Attendance and Achievement in Urban Elementary and Middle Schools: An Instrumental Variables Approach, American Educational Research Journal, 47 (2), 434–465

Gottfried, M.A. (2014) Can Neighbor Attributes Predict School Absences, Urban Education 49 (2) 2016-2050.

Gubbels, J., C. E. van der Put and M. Assink (2019) Risk Factors for School Absenteeism and Dropout: A Meta-Analytic Review, Journal of Youth Adolescence, 48, 1637–1667

Jack, R., C. Halloran, J. Okun, and E. Oster (2023), Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts. American Economic Review: Insights 5 (2): 173–90

Jaume, D. and Willén A. (2019). The Long-Run Effects of Teacher Strikes: Evidence from Argentina. Journal of Labor Economics 37(4), 1097-1139

Johnson, D. (2009). Collateral Damage: The Impact of Work Stoppages on Student Performance in Ontario (e-brief 74). C.D. Howe Institute. Retrieved from http://search.proquest.com/docview/1698040518/

Johnson, D. R. (2011). Do Strikes and Work-to-Rule Campaigns Change Elementary School Assessment Results? Canadian Public Policy / Analyse de Politiques, 37(4), 479–494.

Lavy, V. (2015). Do Differences in School's Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. Economic Journal, 125(88), 397-424.

Liu, J., Lee M. and S. Gershenson (2021). "The short- and long-run impacts of secondary school absences" Journal of Public Economics, 199 104441

McDonald B., Lester, K.J., and Michelson, D. (2022) 'She didn't know how to go back': School attendance problems in the context of the COVID-19 pandemic – A multiple stakeholder qualitative study with parents and professionals, British Journal of Educational Psychology, 93 (1) March 2023 386-401.

Milanovic, K., Blainey, K., Hannay, T., and Wood, C., (2023). The longer-term impact of COVID-19 on pupil attainment and wellbeing. https://www.nuffieldfoundation.org/wp-content/uploads/2022/02/The-longer-term-impact-of-COVID-19-on-pupil-attainment-and-well-being.pdf

Montebruno, P. (2020). Disrupted Schooling: Impacts on achievement from the Chilean school occupations, CEP Discussion Paper No 1696: , Centre for Economic Performance, London School of Economics and Political Science.

Gottfried, M.A. (2014) Can Neighbor Attributes Predict School Absences, Urban Education 49 (2) 2016-2050.

Paris, F. and S. Mervosh (2024) Why School Absences Have 'Exploded' Almost Everywhere, New York Times, March 2024: https://www.nytimes.com/interactive/2024/03/29/us/chronic-absences.html

Patrinos, H., Vegas, E., and Carter-Rau, R. (2022). An Analysis of COVID-19 Student Learning Loss. Policy Research Working Paper 10033. World Bank Group https://documents1.worldbank.org/curated/en/099720405042223104/pdf/IDU00f3f0ca808cde0497e0b88c01fa07f15bef0.pdf

Pischke, J-S (2007) The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years. Economic Journal, 117, 1216-1242

Renaissance Learning & Education Policy Institute (2021), 'Understanding Progress in the 2020/21 Academic Year: Findings from the Summer Term and Summary of All Previous Findings'.

Rose, S., Twist, L., Lord, P., Rutt, S., Badr, K., Hope, C., and Styles, B. (2021), 'Impact of School Closures and Subsequent Support Strategies on Attainment and Socio-Emotional Wellbeing in Key Stage 1', National Foundation for Educational Research, Interim Paper 1.

Ross, S., S. Jianga, M. F. Young, J. L. Meyer and Michael J. Strambler (2023) In-Person Learning During the Pandemic: Student Take-up and School-Level Effects of Remote and Hybrid Instruction on Student Outcomes, unpublished manuscript

Sosu, E.M, S. Dare, C. Goodfellow and M. Klein (2021) Socioeconomic status and school absenteeism: A systematic review and narrative synthesis, BERA Review of Education 9 (3) e3291

Sutton Trust (2021), 'Remote Learning: The Digital Divide'.

Taylor, L. (2023) Prolonged pandemic impact: School absenteeism at record levels, The Brussels Times, 8th Feb 2023 https://www.brusselstimes.com/366824/prolonged-pandemic-impact-school-absenteeism-at-record-levels

Wallace-Wells, D. (2024) Why Children Are Missing More School Now, New York Times, 5th June 2024: https://www.nytimes.com/2024/06/05/opinion/covid-school-attendance-pandemic-closings.html?

Windmeijer, F. (2019). Two-stage Least Squares as Minimum Distance, The Econometrics Journal, 22(1): 1–9

8. Tables and Figures

Figure 1: Weekly calendar of England's pandemic regulations and school restrictions

March 2020		
02	Warnings	Open
09	Warnings	Open
16	Reduce activities	Open
23	1st National lockdown	Key worker children only
30	1st National lockdown	Key worker children only

April 2020		
06	1st National lockdown	Easter holiday
13	1st National lockdown	Easter holiday
20	1st National lockdown	Key worker children only
27	1st National lockdown	Key worker children only

May 2020		
04	1st National lockdown	Key worker children only
11	1st National lockdown	Key worker children only
18	1st National lockdown	Key worker children only
25	1st National lockdown	Key worker children only

June 2020		
01	Phased re-opening	Phased reopening (prim)
08	Phased re-opening	Phased reopening
15	Phased re-opening	Phased reopening (sec)
22	Phased re-opening	Phased reopening
29	Phased re-opening	Phased reopening

July 2020		
06	Reopening, local lockdowns	Phased school reopening
13	Reopening, local lockdowns	Phased school reopening
20	Reopening, local lockdowns	Phased school reopening
27	Reopening, local lockdowns	Summer holiday

August 2020		
03	Reopening, local lockdowns	Summer holiday
10	Reopening, local lockdowns	Summer holiday
17	Reopening, local lockdowns	Summer holiday
24	Reopening, local lockdowns	Summer holiday
31	Reopening, local lockdowns	Summer holiday

September 2020		
07	Reopening, local lockdowns	Open with restrictions
14	Local lockdowns, rule of 6	Open with restrictions
21	Local lockdowns, wfh	Open with restrictions
28	Local lockdowns, wfh	Open with restrictions

October 2020		
05	Local lockdowns, wfh	Open with restrictions
12	1st Tier regs (3 Tiers)	Open with restrictions
19	1st Tier regs (3 Tiers)	Open with restrictions
26	1st Tier regs (3 Tiers)	Half term

November 2020		
02	2nd National Lockdown 5/11	Open with restrictions
09	2nd National Lockdown	Open with restrictions
16	2nd National Lockdown	Open with restrictions
23	2nd National Lockdown	Open with restrictions
30	Lockdown ends 2/12	Open with restrictions

December 2020		
07	2nd Tier Regs (3 Tiers)	Open with restrictions
14	2nd Tier Regs (3 Tiers)	Open with restrictions
21	2nd Tier Regs (4 Tiers)	Christmas holiday
28	2nd Tier Regs (4 Tiers)	Christmas holiday

January 2021		
04	3rd National Lockdown 6th	Open with restrictions
11	3rd National Lockdown	Key worker children only
18	3rd National Lockdown	Key worker children only
25	3rd National Lockdown	Key worker children only

February 2021		
01	3rd National Lockdown	Key worker children only
08	3rd National Lockdown	Key worker children only
15	3rd National Lockdown	Key worker children only
22	3rd National Lockdown	Key worker children only

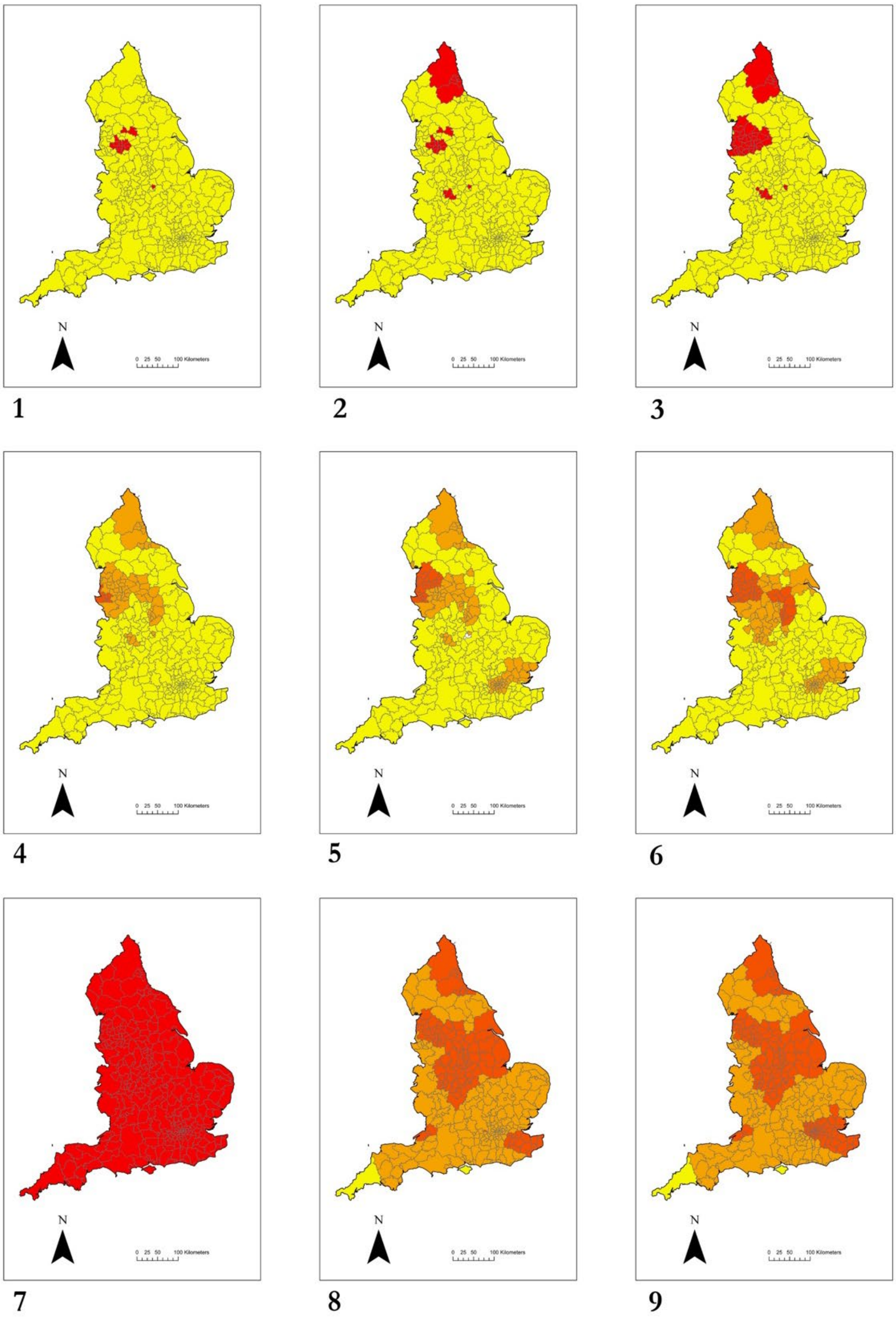
March 2021		
01	3rd National Lockdown	Key worker children only
08	Re-opening step 1, stay home	Open with restrictions
15	Re-opening step 1, stay home	Half term
22	Re-opening step 1, stay home	Open with restrictions
29	Re-opening step 1	Open with restrictions

April 2021		
05	Re-opening step 1	Easter holiday
12	Re-opening step 2	Easter holiday
19	Re-opening step 2	Open with restrictions
26	Re-opening step 2	Open with restrictions

May 2021		
03	Re-opening step 2	Open with restrictions
10	Re-opening step 2	Open with restrictions
17	Re-opening step 3	Open with restrictions
24	Re-opening step 3	Open with restrictions
29	Re-opening step 3	Open with restrictions

Calendar shows general restrictions in orange and school restrictions in blue. School holidays are approximate as these vary from school to school.

Figure 2: Geographical distribution of lockdowns and Tier regulations in England, September to December 2020.



Notes: Panels 1-3, Local Lockdowns 1st September 2020 to 14th October; Panels 4-6 the First Tier regulations (from the 14th October 2020 to 5h November 20020); Panel 7 Second National Lockdown (from 5th November to 2nd December 2020; Panel 8 and 9, initial part of the All Tiers regulations (in our sample, it spans from 2nd December to 18th December or the last term day). We categorise Local Lockdowns as Tier 4 and periods without restrictions or with Tier 1 restrictions as Tier 1.
Key: Red - Tier 4 or lockdown; Dark Orange – Tier 3; Light Orange – Tier 2; Yellow – Tier 1/no restrictions

Figure 3: Cohorts in our dataset

Year	Term	Cohort 4 Yeargroup	Cohort 5 Yeargroup	Cohort 6 Yeargroup	Cohort 7 Yeargroup	Cohort 8 Yeargroup	Cohort 9 Yeargroup	Cohort 10 Yeargroup	Cohort 11 Yeargroup	Cohort 12 Yeargroup	Cohort 13 Yeargroup	Pandemic	Data
2017/18	Autumn	1	2	3	4	5	6	7	8	9	10	Pre	KS and ab
2017/18	Spring	1	2	3	4	5	6	7	8	9	10	Pre	KS and ab
2017/18	Summer	1	2	3	4	5	6	7	8	9	10	Pre	KS and ab
2018/19	Autumn	2	3	4	5	6	7	8	9	10	11	Pre	KS and ab
2018/19	Spring	2	3	4	5	6	7	8	9	10	11	Pre	KS and ab
2018/19	Summer	2	3	4	5	6	7	8	9	10	11	Pre	KS and ab
2019/20	Autumn	3	4	5	6	7	8	9	10	11		Pre	No data
2019/20	Spring	3	4	5	6	7	8	9	10	11		Post	No data
2019/20	Summer	3	4	5	6	7	8	9	10	11		Post	No data
2020/21	Autumn	4	5	6	7	8	9	10	11			Post	No KS dat
2020/21	Spring	4	5	6	7	8	9	10	11			Post	No KS dat
2020/21	Summer	4	5	6	7	8	9	10	11			Post	No KS dat
2021/22	Autumn	5	6	7	8	9	10	11				Post	KS and ab
2021/22	Spring	5	6	7	8	9	10	11				Post	KS and ab
2021/22	Summer	5	6	7	8	9	10	11				Post	KS and ab

Table 1: Absence autumn term, pre and post pandemic

	2018/19		2021/22	
	Mean	s.d.	Mean	s.d.
	3.526	5.325	9.268	11.377
Unforced absence	3.526	5.325	7.691	10.753
Forced absence (code X)	0	0	1.763	3.627
Last year's absence	3.661	5.113	13.507	15.161
Last year's unforced absence	0	0	5.613	10.227
Last year's forced absence (code X)	0	0	8.672	11.697
Observations	3772718		3754898	

Notes: Unweighted means and standard deviations. Forced absence (code X) refers to absence caused by self-isolation due to COVID-19 infection in a student's household or support group.

Table 2 Persistence of absence over two years, autumn 2020/21-2021/22, and autumn 2017/18-2018/19, Year Groups 1-7 in 2017/18

	Sessions absent % OLS	Sessions absent % OLS	Sessions absent % OLS	Sessions absent % 2SLS	Sessions absent % 2SLS	Sessions absent % 2SLS
Autumn absence last year	0.306***	0.299***	0.183***	0.304***	0.418***	0.654***
	(0.007)	(0.007)	(0.007)	(0.026)	(0.032)	(0.067)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pupil fixed effects	No	No	Yes	No	No	Yes
Pandemic dummy	Yes	Yes	Yes	Yes	Yes	Yes
Cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pupil characteristics	No	Yes	Yes	No	Yes	Yes
COVID deaths and case rate	No	Yes	Yes	No	Yes	Yes
Unemployment rate	No	Yes	Yes	No	Yes	Yes
F-test excluded instr.	-	-	-	72.236	51.871	20.277
Over id test p-value	-	-	-	0.0744	0.0739	0.2779
R squared	0.259	0.276	0.690	-	-	-
Observations	7527570	7527570	7509702	7527570	7527570	7509702

Notes: Table reports regression coefficients and standard errors. Significance ***1%, **5%, *10%. Standard errors clustered at LEA level. Pupil characteristics include gender, ethnic group dummies (7 categories), free meals eligible, English first language, month of birth dummies, and MSOA Index of Deprivation Affecting Children. COVID variables are COVID death rate and case rate (zero for 2017/18-2018/19) at school MSOA level. Unemployment claimant rate is at school LSOA level. Data is for pupil in years and cohorts set out in Figure 3. All regressions include a pre/post pandemic year dummy (2020/21=1). In columns 4-6, previous year's absence is predicted from COVID policy variable instruments. The instruments in columns 4-6 are number of days in each Tier in autumn term 2020 and indicator of whether LA advised schools to follow government reopening policy in Summer Term 2020, interacted with pupil home LSOA IDACI score. 2SLS regressions in columns 4-6 include controls for baseline main effects of these policy variables (i.e., un-interacted with IDACI).

Table 3: First stage regressions corresponding to 2SLS regressions in Table 2, and Tier 4 placebo test

Autumn absence	Main effects of policies	Excluded instruments: Policy x home IDACI score	Main effects of policies (Tier 4 placebo test)	Excluded instruments: Policy x home IDACI score (Tier 4 placebo test)
Christmas and New Year 2020 (placebo)				
Tier 4, 2nd tier regs days	-	-	-0.062	-0.172*
			(0.027)	(0.096)
Autumn 2020				
Tier 3, 2nd tier regs days	0.144***	0.437***	0.127***	0.532***
	(0.049)	(0.083)	(0.048)	(0.097)
Tier 2, 2nd tier regs days	0.062	0.439***	-0.028	0.577***
	(0.042)	(0.065)	(0.042)	(0.097)
Tier 3, 1st tier regs days	-0.077	0.045	-0.060	0.006
	(0.053)	(0.106)	(0.056)	(0.105)
Tier 2, 1st tier regs days	-0.007	-0.090	-0.007	-0.084
	(0.028)	(0.067)	(0.024)	(0.076)
Local Lockdown days	0.021	0.002	0.025	0.005
	(0.023)	(0.053)	(0.005)	(0.048)
Summer 2020				
LA follows govt	-0.505	-0.555	0.192	-0.600
	(0.459)	(1.246)	(0.021)	(1.331)
School fixed effects		Yes		Yes
Pupil fixed effects		Yes		Yes
Pandemic dummy		Yes		Yes
Cohort fixed effects		Yes		Yes
Pupil characteristics		Yes		Yes
COVID variables		Yes		Yes
Unemployment rate		Yes		Yes
F-test excluded instr.		20.277		12.750
F-test placebo treatment		-		3.227
R squared		0.700		0.700
Observations		7509702		7509702

Notes: See Table 2

Table 4: Persistence of absence over two years, differences by student group

	Sessions absent % Low depriv. 2SLS	Sessions absent % High depriv. 2SLS	Sessions absent % Not free meals 2SLS	Sessions absent % Free meals 2SLS	Sessions absent % North 2SLS	Sessions absent % South 2SLS	Sessions absent % Conurbation 2SLS	Sessions absent % Urban 2SLS	Sessions absent % Rural 2SLS	Sessions absent % Male 2SLS	Sessions absent % Female 2SLS	Sessions absent % Second. 2SLS	Sessions absent % Primary 2SLS
Autumn abs. last year	0.537*** (0.068)	0.557*** (0.056)	0.228*** (0.044)	0.457*** (0.041)	0.508*** (0.036)	0.501*** (0.048)	0.455*** (0.046)	0.617*** (0.048)	0.590*** (0.134)	0.487*** (0.030)	0.527 (0.034)	0.715*** (0.039)	0.284*** (0.030)
School fix. eff.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pandemic dum.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort fix. eff.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil Xs.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
COVID Xs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unemployment	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F excl. instr.	16.409	17.934	31.080	20.573	47.226	75.443	30.396	52.329	21.009	71.237	67.835	61.564	52.961
Observations	3765279	3761562	6170238	1355463	3657916	3869174	2849279	3545536	1129035	3843670	3683701	4172688	3354827

Notes: Specifications similar to Table 2, column 5. See Table 2 for details.

Table 5: Effect of pandemic absence on chronic absence

Autumn absence	More than 10% missed	More than 20% missed	More than 40% missed
Autumn absence last year	0.016***	0.026***	0.008***
	(0.002)	(0.003)	(0.001)
Mean pre-pandemic	0.080	0.014	0.002
Mean post-pandemic	0.341	0.103	0.023
School fixed effects	Yes	Yes	Yes
Pupil fixed effects	Yes	Yes	Yes
Pandemic dummy	Yes	Yes	Yes
Cohort fixed effects	Yes	Yes	Yes
Pupil characteristics	Yes	Yes	Yes
COVID variables	Yes	Yes	Yes
Unemployment rate	Yes	Yes	Yes
F-test excluded instr.	20.277	20.277	20.277
Over id test p-value	0.137	0.110	0.170
Observations	7509702	7509702	7509702

Notes: Specifications similar to Table 2, column 6, replacing absence rates with indicators of absence over thresholds specified in column headings.

Table 6: Effect of compulsory and non-compulsory absence on subsequent absence

Autumn absence previous year	OLS	2SLS
Non-compulsory absence	0.361***	0.563***
	(0.009)	(0.032)
Compulsory absence (Code X)	-0.007***	-0.053***
	(0.002)	(0.007)
School fixed effects	Yes	Yes
Pupil fixed effects	Yes	Yes
Pandemic dummy	Yes	Yes
Cohort fixed effects	Yes	Yes
Pupil characteristics	Yes	Yes
COVID variables	Yes	Yes
Unemployment rate	Yes	Yes
F-test excluded instr.	-	168.593
Over id test p-value	-	0.059
Observations	7509702	7509702

Notes: See Table 2. Forced absence (code X) refers to absence caused by self-isolation due to COVID-19 infection in a student's household or support group. Column 2 predicts unforced absence from autumn 2020 local COVID policies.

Table 7: Regression estimates of pandemic-induced absence on achievement. Influence of autumn absence on primary school students' KS2 achievement (absence in 2020/21, 2017/18; KS4 in 2021/22, 2018/19)

	KS2 maths percentile OLS	KS2 grammar, spelling, punctuation percentile OLS	KS2 reading percentile OLS	KS2 maths percentile 2SLS	KS2 grammar, spelling, punctuation percentile 2SLS	KS2 reading percentile 2SLS	KS2 maths percentile 2SLS	KS2 grammar, spelling, punctuation percentile 2SLS	KS2 reading percentile 2SLS
Autumn percent sessions pupil absent	-0.179*** (0.004)	-0.123*** (0.003)	-0.064*** (0.003)	-0.997*** (0.144)	-1.058*** (0.133)	-0.344*** (0.101)	-1.120*** (0.178)	-1.253*** (0.179)	-0.317*** (0.116)
School fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
School x IDACI	No	No	No	No	No	No	Yes	Yes	Yes
Pupil KS1 scores	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pupil characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
COVID variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unemployment rate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LA maint. x pandemic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-test excluded instr.				26.302	26.302	26.302	22.436	22.436	22.436
R squared	0.548	0.583	0.464						
Observations	1163688	1163688	1163688	1163688	1163688	1163688	1163688	1163688	1163688

Notes: Table reports regression coefficients and standard errors. Significance ***1%, **5%, *10%. Standard errors clustered at LEA level. Pupil KS1 scores are age 7 percentile scores in reading, writing and maths. Pupil characteristics include gender, ethnic group dummies (7 categories), free meals eligible, English first language, month of birth dummies, and MSOA Index of Deprivation Affecting Children. COVID variables are COVID death rate and case rate (zero for 2018/19) at school MSOA level. Unemployment claimant rate is at school LSOA level. Data is for pupils taking KS2 in 2018/19 and 2021/22, with absence recorded in autumn 2017/18 and 2020/21 respectively, when pupils were in Year 5. All regressions include a pre/post pandemic year dummy (2020/21=1). Instruments in columns 4-9 are number of days in each Tier in autumn term 2020 and indicator of whether LA advised schools to follow government reopening policy in Summer Term 2020, interacted with pupil home LSOA IDACI score. 2SLS regressions in columns 4-9 include controls for baseline main effects of these policy variables (i.e., un-interacted with IDACI).

Table 8: First stage regressions corresponding to 2SLS regressions in Table 7, and Tier 4 placebo test

KS2 autumn Year 5 percent sessions absent	Main effects of policies	Excluded instruments: Policy x home IDACI score	Main effects of policies (Tier 4 placebo test)	Excluded instruments: Policy x home IDACI score (Tier 4 placebo test)
Christmas and New Year 2020 (placebo)				
Tier 4, 2nd tier regs days	-	-	0.011	-0.066
			(0.018)	(0.084)
Autumn 2020				
Tier 3, 2nd tier regs days	0.048	0.392***	0.043	0.424***
	(0.029)	(0.085)	(0.030)	(0.087)
Tier 2, 2nd tier regs days	-0.004	0.329***	-0.013	0.382***
	(0.022)	(0.071)	(0.024)	(0.080)
Tier 3, 1st tier regs days	-0.065**	0.087	-0.062**	0.069
	(0.027)	(0.086)	(0.029)	(0.087)
Tier 2, 1st tier regs days	0.003	-0.073	0.003	-0.070
	(0.016)	(0.073)	(0.016)	(0.077)
Local Lockdown days	0.016	-0.003	0.017	-0.004
	(0.012)	(0.039)	(0.012)	(0.038)
Summer 2020				
LA follows govt	-0.181	-1.169	-0.174	-1.200
	(0.262)	(1.236)	(0.265)	(1.260)
School x IDACI		Yes		Yes
Pupil KS1 scores		Yes		Yes
Pupil characteristics		Yes		Yes
COVID variables		Yes		Yes
Unemployment rate		Yes		Yes
LA maintained x pandemic		Yes		Yes
F-test excluded instr.		22.436		16.828
F-test placebo treatment		-		0.618
R squared		0.287		0.287
Observations		1163688		1163688

Notes: see Table 7

Table 9: Regression estimates of effects of pandemic-induced absence on KS2 achievement; differences by student group

	Low deprivation 2SLS	High deprivation 2SLS	Not free meals 2SLS	Free meals 2SLS	North 2SLS	South 2SLS	Conurbation 2SLS	Urban 2SLS	Rural 2SLS	Male 2SLS	Female 2SLS
KS2 Maths	-1.435***	-1.095***	-1.759***	-0.918***	-1.099***	-1.082***	-0.589***	-1.031***	-0.782*	-0.924***	-1.233***
	(0.465)	(0.274)	(0.361)	(0.182)	(0.195)	(0.399)	(0.163)	(0.279)	(0.414)	(0.191)	(0.183)
KS2 GPS	-1.813***	-1.149***	-1.885***	-0.917***	-1.154***	-1.144***	-0.718***	-1.144***	-1.076***	-1.315***	-1.144***
	(0.506)	(0.265)	(0.366)	(0.170)	(0.189)	(0.371)	(0.164)	(0.263)	(0.401)	(0.202)	(0.182)
KS2 Reading	0.045	-0.782***	-0.409*	-0.389***	-0.502***	-0.407**	-0.018	-0.331*	-0.281	-0.520***	-0.124
	(0.351)	(0.224)	(0.231)	(0.144)	(0.133)	(0.197)	(0.167)	(0.178)	(0.305)	(0.130)	(0.142)
F-test excluded instr.	6.626	7.762	9.109	11.924	19.907	17.888	15.744	11.048	9.137	17.932	20.980
Observations	584069	578223	948084	213207	563859	599642	442642	543584	174050	587115	576432

Notes: All regressions include controls and fixed effects listed in Table 7. Specification based on columns 7-9 of Table 7.

Table 10: Effect of compulsory and non-compulsory absence on subsequent KS2 achievement

	OLS			2SLS		
	Maths	GPS	Reading	Maths	GPS	Reading
Non-compulsory absence	-0.210***	-0.127***	0.005	-5.542***	-6.172***	-1.830***
	(0.004)	(0.004)	(0.004)	(0.850)	(0.846)	(0.609)
Compulsory absence	-0.037***	-0.030***	-0.038***	1.022***	1.172***	0.326***
	(0.006)	(0.005)	(0.005)	(0.169)	(0.171)	(0.121)
School x IDACI effects	Yes	Yes	Yes	Yes	Yes	Yes
Pandemic dummy	Yes	Yes	Yes	Yes	Yes	Yes
Cohort fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Pupil characteristics	Yes	Yes	Yes	Yes	Yes	Yes
COVID variables	Yes	Yes	Yes	Yes	Yes	Yes
Unemployment rate	Yes	Yes	Yes	Yes	Yes	Yes
F-test excluded instr.	-	-	-	14.796	24.796	14.796
Over id test p-value	-	-	-	0.170	0.373	0.015
Observations	1163688	1163688	1163688	1163688	1163688	1163688

Notes: See Table 7. Forced absence (code X) refers to absence caused by self-isolation due to COVID-19 infection in a student's household or support group. Column 2 predicts unforced absence from autumn 2020 local COVID policies.

Table 11: Regression estimates of effect of pandemic-induced absence on achievement. Influence of autumn absence on secondary school students' KS4 achievement (absence in 2020/21, 2017/18; KS4 in 2021/22, 2018/19)

	Total KS4 points OLS	Total KS4 points 2SLS	Total KS4 points 2SLS
Autumn percent sessions pupil absent	-0.311***	0.031	0.139
	(0.007)	(0.068)	(0.087)
School fixed effect	Yes	Yes	No
School x IDACI	No	No	Yes
Pupil KS1& KS2	Yes	Yes	Yes
Pupil characteristics	Yes	Yes	Yes
COVID variables	Yes	Yes	Yes
Unemployment rate	Yes	Yes	Yes
LA maint. x pandemic	Yes	Yes	Yes
F-test excluded instr.		12.988	11.592
R squared	0.617		
Observations	1036237	1036237	1036237

Notes: Table reports regression coefficients and standard errors. Significance ***1%, **5%, *10%. Standard errors clustered at LEA level. Dependent variable is Key Stage 4 total points converted to percentiles of the distribution across pupils in a given year. Pupil KS1/KS2 scores are age 7/11 percentile scores in reading, writing and maths. Pupil characteristics include gender, ethnic group dummies (7 categories), free meals eligible, English first language, month of birth dummies, and MSOA Index of Deprivation Affecting Children. COVID variables are COVID death rate and case rate (zero for 2018/19) at school MSOA level. Unemployment claimant rate is at school LSOA level. Data is for pupils taking KS4 in 2018/19 and 2021/22, with absence recorded in autumn 2017/18 and 2020/21 respectively, when pupils were in Year 10. All regressions include a pre/post pandemic year dummy (2020/21=1). Instruments in columns 2-3 are number of days in each Tier in autumn term 2020 and indicator of whether LA advised schools to follow government reopening policy in Summer Term 2020, interacted with pupil home LSOA IDACI score. 2SLS regressions in columns 2-3 include controls for baseline main effects of these policy variables (i.e., un-interacted with IDACI).

Table 12: Descriptive statistics (post pandemic period)

	Absence		KS2	
	Mean	sd	Mean	sd
Autumn absence (percentage)	9.268	11.377	-	-
Autumn absence lag	13.507	15.163	10.197	12.431
Maths	-	-	50.182	28.525
GPS	-	-	49.924	28.483
Reading	-	-	49.168	28.170
KS1 reading	-	-	34.314	25.246
KS1 writing	-	-	32.947	24.811
KS1 maths	-	-	32.651	25.756
IDACI home	0.176	0.123	0.177	0.123
Claimant count percent	42.600	25.602	43.120	25.899
FSM eligible	0.217		0.225	
Female	0.489		0.505	
White	0.739		0.739	
Other ethnic group	0.018		0.021	
Black	0.108		0.115	
Asian	0.056		0.057	
Chinese	0.004		0.005	
Mixed	0.062		0.064	
Uncertain ethnic group	0.014		0.009	
English first language	0.826		0.790	
Language uncertain	0.004		0.001	
Born September	0.086		0.087	
Born October	0.086		0.090	
Born November	0.082		0.083	
Born December	0.083		0.084	
Born January	0.083		0.084	
Born February	0.076		0.075	
Born March	0.083		0.084	
Born April	0.080		0.079	
Born May	0.085		0.083	
Born June	0.083		0.083	
Born July	0.087		0.086	
Born August	0.086		0.084	
LA maintained	-		0.590	
Policy variables for 2020/21				
Follow government	0.277		0.273	
Local lockdown (days)	7.393	13.700	7.435	13.785
Tier 1, 1st tier regs (days)	10.571	10.116	10.535	10.106
Tier 2, 1st tier regs (days)	9.019	9.197	9.079	9.198
Tier 3, 1st tier regs (days)	2.410	5.763	2.386	5.725
Tier 1, 2nd tier regs (days)	0.205	1.854	0.199	1.827
Tier 2, 2nd tier regs (days)	8.622	8.028	8.632	8.024
Tier 3, 2nd tier regs (days)	8.173	8.022	8.170	8.019
COVID death rate to 04/21	0.183	0.098	0.183	0.098
COVID case rate 2020 100k	3.075	1.569	3.079	1.574
COVID case rate 2021 100k	8.558	1.963	3.723	0.915

Notes: Unweighted means and standard deviations. Pupil data.

Table 13: Regressions of effect of 2020 policies on absence in 2021/22

	Maths		GPS		Reading	
KS2 2022 scores	Main effects of policies	Excluded instruments : Policy x home IDACI score	Main effects of policies	Excluded instruments : Policy x home IDACI score	Main effects of policies	Excluded instruments : Policy x home IDACI score
Autumn 2020						
Tier 3, 2nd tier regs days	0.005	-0.275***	-0.042	-0.319***	0.029	-0.050
	(0.030)	(0.079)	(0.036)	(0.068)	(0.033)	(0.065)
Tier 2, 2nd tier regs days	0.070***	-0.475***	0.031	-0.460***	0.065**	-0.183***
	(0.025)	(0.075)	(0.034)	(0.066)	(0.028)	(0.062)
Tier 3, 1st tier regs days	0.012	-0.080	-0.014	-0.037	0.057**	-0.086
	(0.029)	(0.077)	(0.031)	(0.067)	(0.026)	(0.066)
Tier 2, 1st tier regs days	0.014	0.000	-0.012	0.056	0.021	0.006
	(0.022)	(0.068)	(0.021)	(0.062)	(0.020)	(0.061)
Local Lockdown days	-0.001	0.054	0.012	0.005	0.012	0.034
	(0.016)	(0.041)	(0.015)	(0.037)	(0.014)	(0.040)
Summer 2020						
LA follows govt	0.626	0.293	0.412	-0.351	0.405	-1.091
	(0.392)	(1.354)	(0.381)	(1.198)	(0.334)	(1.261)
School fixed effects		Yes		Yes		
Pupil KS1 scores		Yes		Yes		
Pupil characteristics		Yes		Yes		
COVID variables		Yes		Yes		
Unemployment rate		Yes		Yes		
LA maintained x pandemic		Yes		Yes		
F-test policy x IDACI		16.466		22.477		5.160
F-test p-value		0.0000		0.0000		0.0001
R squared		0.546		0.582		0.464
Observations		1163688		1163688		1163688

Notes: See Table 7. Regressions of KS2 outcomes on COVID-19 policy variables.

Table 14: Effect of characteristics on absence, OLS regression Table 2, column 2

Autumn absence		
	Coefficient	Standard error
Autumn absence previous year	0.299***	(0.007)
Pandemic dummy (for 2020/21-2021/22)	2.987***	(0.281)
IDACI score home	3.646***	(0.151)
COVID death rate	0.258	(0.204)
Case rate 2020	-0.745***	(0.043)
Case rate 2021	0.182***	(0.024)
Claimant count rate	-0.004***	(0.001)
Female	0.096***	(0.013)
Other ethnic group	-1.287***	(0.073)
Black	-0.861***	(0.043)
Asian	-2.066***	(0.063)
Chinese	-2.162***	(0.051)
Mixed	-0.337***	(0.026)
Uncertain	-0.099	(0.064)
FSM eligible	2.273***	(0.050)
English not first language	-0.517***	(0.026)
Language uncertain	0.366***	(0.081)
Born October	-0.053***	(0.014)
Born November	-0.042***	(0.014)
Born December	-0.036**	(0.015)
Born January	-0.100***	(0.015)
Born February	-0.094***	(0.015)
Born March	-0.151***	(0.015)
Born April	-0.171***	(0.016)
Born May	-0.183***	(0.014)
Born June	-0.206***	(0.015)
Born July	-0.207***	(0.013)
Born August	-0.218***	(0.014)
R squared	0.276	
Observations	752750	

Notes: OLS school fixed effects regression results for Table 2, column 2.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

2031	Stephen Gibbons Sandra McNally Piero Montebruno	Pupil absence and the pandemic
2030	Johannes Boehm Thomas Chaney	Trade and the end of antiquity
2029	Jay Euijung Lee Martina Zanella	Learning about women's competence: The dynamic response of political parties to gender quotes in South Korea
2028	Christos Genakos Blair Yuan Lyu Mario Pagliero	Asymmetric pass-through and competition
2027	Fabrice Defever Alejandro Riaño Gonzalo Varela	Evaluating the impact of export finance support on firm-level export performance: Evidence from Pakistan
2026	Andrés Barrios-Fernández Christopher Neilson Seth Zimmerman	Elite universities and the intergenerational transmission of human and social capital
2025	Agnes Norris Keiller	Brexit and investment
2024	Lídia Farré Libertad González Claudia Hupkau Jenifer Ruiz-Valenzuela	Paternity leave and child development
2023	Gabriel M. Ahlfeldt Nikodem Szumilo Jagdish Tripathy	Housing-consumption channel of mortgage demand
2022	Agnes Norris Keiller Tim Obermeier Andreas Teichgraeber John Van Reenen	When trade drives markup divergence. An application to auto markets

2021	Amanda Dahlstrand Nestor Le Nestour Guy Michaels	Online versus in-person services: Effects on patients and providers
2020	Hanwei Huang Kalina Manova Oscar Perelló Frank Pisch	Firm heterogeneity and imperfect competition in global production networks
2019	Jan David Bakker Nikhil Datta	Avenging the tenants: Regulating the middle man's rents
2018	Agnes Norris Keiller Áureo de Paula John Van Reenen	Production function estimation using subjective expectations data
2017	Philippe Aghion Lint Barrage David Hémous Ernest Liu	Transition to green technology along the chain supply
2016	Tom Schmitz Italo Colantone Gianmarco Ottaviano	Regional and aggregate economic consequences of environmental policy
2015	Agnes Norris Keiller Tim Obermeier Andreas Teichgraeber John Van Reenen	An engine of (pay) growth? Productivity and wages in the UK auto industry
2014	Antonin Bergeaud Pierre Cahuc Clément Malgouyres Sara Signorelli Thomas Zuber	The wage of temporary agency workers
2013	Richard Hornbeck Guy Michaels Ferdinand Rauch	Identifying agglomeration shadows: Long-run evidence from ancient ports

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE