Can less be more? Instruction time and attainment in English secondary schools: Evidence from panel data

Vaughan Connolly ⓘD

How to cite this article

Peer review

This article has been peer-reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymized during review.

Copyright

Open access

The *London Review of Education* is a peer-reviewed open-access journal.

LONDON
REVIEW OF
EDUCATION

Can less be more? Instruction time and attainment in English secondary schools: Evidence from panel data

Vaughan Connolly* – *University of Cambridge, UK*

Abstract

How to best utilize curriculum time has long been a question for England's schools, which are free to vary time between subjects or to extend the school day/week. This question has now risen to national prominence as policymakers consider ways to help support students catch up lost learning after the COVID-19 pandemic. This article explores the relationship between instruction time on General Certificate of Secondary Education (GCSE) attainment and value-added scores for 2,815 English secondary schools using panel data from the 2010–14 School Workforce Census. Robust estimates suggest that instruction time has a small effect on attainment in each of English, mathematics, science and humanities. Based on sub-group analysis, effects sizes are also estimated for students with varying prior achievement and for those in receipt of free school meals. These too appear small, suggesting, vis-à-vis instructional time, sometimes less might be more.

Keywords: instruction time, time use, workload, secondary schools

Introduction

How best to allocate curriculum time is a critical issue for England's secondary schools, which are free to vary the amount of time spent teaching each subject in each year group. This is particularly true for academy schools, which are permitted to alter the length of their school day, week and year. Traditionally, discussions have focused on whether to give additional time to core subjects or whether students need a longer school week to reach desired progress levels. However, as COVID-19 impacts schools, the issue of instructional time appears germane for two new reasons. First, should head teachers worry about reducing instructional hours to meet social distancing and other safety requirements? Second, might extending the school day/year help students 'catch up' learning lost to the pandemic? While international research has suggested small returns to additional instruction time (Bietenbeck and Collins, 2020; Bingley et al., 2018; Lavy, 2015), prior research has been inconclusive (Hanushek, 2015). Looking to cross-country evidence fails to show a clear pattern (Van Damme, 2014), and in the face of vast differences in countries' allocation of instruction time, the issue has become contentious and politicized (Berliner and Biddle, 1995; Patall et al., 2010). This international uncertainty highlights the importance of directly studying England's secondary schools, as 'even the best identified' study from other countries 'presents challenges when it comes to generalising' (Hanushek, 2015: F395). As this has yet to be done (EEF, 2019), this article addresses an important evidence gap.

*Email: vnc20@cam.ac.uk00

To address this evidence gap, administrative data from England's School Workforce Census (SWC) are analysed for 2,815 secondary schools from 2010 to 2014. Panel methods are used to explore the relationship between instruction time in English, mathematics, science and humanities departments and value-added performance at General Certificate of Secondary Education (GCSE) level, finding effect size estimates of 0.08, 0.09, 0.07 and 0.07 standard deviations (SD), respectively. I also ask 'what is the possible effect of an extra hour of timetabled instruction, per Year 11 class, in each subject'? To investigate important equity issues, this analysis is extended to student sub-groups according to prior academic background and free school meals (FSM) status. Finally, I look beyond the core subjects and explore the link between overall time per student in Key Stage (KS) 4 (Years 10 and 11) and the percentage of students gaining five GCSEs at grade C or better.

Specifically, my research questions are:

1. What is the relationship between GCSE value-added performance and instruction time in English, science, mathematics and humanities departments? To what extent does this association vary when analysed by subject area?
2. What is the relationship between GCSE value-added performance and increased instruction time at KS4 in English, science, mathematics and humanities for students grouped by prior attainment and FSM status?

Previous meta-analyses

Although not yet empirically investigated in England's secondary schools, the issue of 'time in schools' has long been debated, particularly in the USA, and has led to a widely held claim: learning time is an important, but not sufficient, predictor of student progress. As an 'empty vessel', what people do with their time is as important as actually having time (Fraser et al., 1987; Karweit, 1984; Scheerens et al., 2007). In a recent update, Scheerens et al. (2013) provide an excellent 'state of the art' review alongside a meta-analysis of studies between 1985 and 2011. First reviewing early meta-analyses, they find effects sizes ranging from (Cohen's) d=0.31 to d=0.52 (average d=0.37) for assessments of 'instructional time' and 'time on task', with higher estimates for the latter. However, these effect sizes are queried for methodological problems, particularly in older articles, a key issue being different operationalizations of time. While the two most common measures are 'instructional time' (for example, timetabled time) and 'time on task' (class-time that is focused on learning activities, earlier known as 'engaged time'), studies have also measured time as 'attention', 'class attendance', 'engagement' or 'classroom management' (Scheerens et al., 2013). Additionally, earlier studies are reliant on cross-sectional data, small sample sizes (particularly when measuring time on task) and do not always account for the multi-level structure of data or have weak identification strategies.

Looking to syntheses of meta-analyses, John Hattie's are arguably most well-known. When reporting the effect size of 'time on task', Hattie (2009) first finds d=0.38, which he later updates to 0.62 when including an analysis of time use in higher education (Hattie, 2015). This shift highlights a weakness in the application of meta-analyses, where the combination of results from differing levels of education (primary through tertiary) may influence the outcome. In addition, Scheerens et al. (2013) note that the articles Hattie draws upon exhibit methodological weaknesses as listed above. In their updated meta-analysis, Scheerens et al. (2013) examine articles published between 2005 and 2011, and find 12 studies amenable to quantitative analysis. These provide 31 effect size estimates: 10 for language, 18 for mathematics and 3 for science.

No estimates were found for other subjects. With a combined effect size of d=0.10, these studies were considerably lower than the d=0.37 reported for older articles, with the difference attributed to methodological quality (Hanushek, 2015; Scheerens et al., 2013). In this context, high quality is held to be investigations based on longitudinal or panel data, drawn from large-scale studies utilizing standardized outcome measures (Scheerens et al., 2013). Also important are the definitional clarity of time measures (with later articles settling on 'allocated instructional time'), appropriate controls and more rigorous identification strategies, such as using fixed effects (FE) to control for unobserved variable bias. This latter point was ably demonstrated by Lavy (2015), who examined PISA 2006 data and compared ordinary least squares (OLS) with student FE, giving the effects of an additional hour of instruction time as 0.196SD for OLS and 0.058SD for FE estimators.

Recent high-quality studies also suggest smaller effect sizes. In a quasi-experimental analysis of admissions lottery data for New York City charter schools, Dobbie and Fryer Jr (2013) observe an effect of 0.05SD to mathematics attainment via a 25 per cent or greater increase in instruction time. In a study similar to Lavy (2015), Rivkin and Schiman (2015) analyse PISA 2009 data, although instead of applying student FE they use school-by-grade and school-by-subject FE and analyse variation both across subjects (mathematics and language arts) and grades (Years 9 and 10) within each school. Their study reveals robust effects ranging from 0.023 to 0.087SD per one hour increase. They also find evidence for a non-linear relationship reflecting a diminishing return on time, with stronger benefits to additional time in 'high-quality' classroom environments. Two weaknesses of PISA data, however, are their self-report and cross-sectional nature. These were addressed recently by Bingley et al. (2018), who utilized Danish administrative data to examine the effect of accumulated instruction time on ninth grade (age 15) performance in Danish, English and mathematics. By assuming a uniform effect across subjects, and applying student FE to analyse variation across subjects, they observed an effect of approximately 0.06SD (per hour, weekly, over grades 1 to 9).

All told, this brief review highlights the importance of methodological rigour and gives indicators of research quality. These indicators support the research questions and methodology proposed in this article, as described next.

Methodology

Data sources

This study analyses administrative data from 2,815 English secondary schools over five years (2010 to 2014), gathered in the annual SWC, which started in 2010. Analysis over a longer time period is prevented by a change in school performance metrics in 2015. Measures of instruction time are derived from the SWC, and are then matched at the school level to inspection data from the Office for Standards in Education, Children's Services and Skills (Ofsted), school performance tables, and a set of rich controls from data published by the Department for Education (DfE, n.d.).

English secondary schooling is partitioned into KS3, KS4, KS5. These equate to Years 7 to 9, 10 to 11 and 12 to 13, respectively. At the end of Year 11, students are generally aged 16 and typically sit public GCSE examinations. At the time, attainments in core 'academic' subjects are used to calculate value-added performance indicators for each GCSE English Baccalaureate (EBacc) 'pillar' or subject group (using the best qualifying score). It is these indicators that are analysed in this study. These measures solely compare department-level performance at the end of KS4 against students'

overall performance at KS2 in mathematics and English (for example, 'value-added' measures are intended to show a student's 'progress' over a five-year span, with no other adjustments made). School-level value-added is used, as opposed to student-level data, as these indicators have national policy significance. Furthermore, they are methodologically sufficient and publicly available, thus removing the need for access to sensitive data and allowing this study to be completed in a timely fashion.

The DfE (2015) also publishes value-added performance for student sub-groups based on prior KS2 attainment and FSM status. Such sub-groups classify students as having low, average or high prior attainment, depending on whether they achieved level three (or lower), level four or level five (or higher) by the end of KS2 (DfE, 2016). When analysing the relationship between time and attainment, I must assume that schools allocate time evenly across all such groups, even though some schools may provide additional time for students with low prior attainment, for example; unfortunately, the data captured in the census does not identify such practices. In total, there are five dependent variables per subject, or 20 in total. Two additional performance measures are considered in this study, owing to their national policy significance and as a methodological robustness check. The first indicates the percentage of students gaining grade C or better in five or more GCSE subjects or equivalent, while the second requires two of these five subjects to include English and mathematics (hereafter referred to as ptac5 and ptac5em, respectively).

As theory and prior empirical evidence suggest a curvilinear functional form to instructional time, it is essential to analyse outlying values of instructional time before testing for non-linear effects. This prevents erroneous conclusions arising from a handful of extreme values exerting undue influence. Such outliers are seen in fewer than 143 schools per year, which have small KS4 cohorts (for example, closing or opening schools) or have potential data quality issues. On a practical level, schools outside of the 1st and 99th percentile do not reflect the experience seen in other schools; hence, it was deemed suitable to exclude outlying schools from ongoing analysis. Outlying value-added data points and schools showing large swings in hours are also excluded (approximately 65 per year) for the same reasons.

Theoretical model

The Carroll model (1989) suggests that a student's performance is influenced by five factors: Aptitude, Opportunity to Learn, Perseverance, Quality of Instruction and Ability to Understand Instruction. There is also a suggestion that student performance suffers from 'diminishing returns' with regard to time, indicating a non-linear functional form (Scheerens et al., 2013). Combining SWC, School Census and Ofsted data (2010–14), this study approximates the Carroll model by considering performance at the school level, using aggregated student-level data as individual-level data is unavailable. To avoid the ecological fallacy, it is essential to recall that what is true at the group level may not necessarily be correct at the individual level. However, it is still valid to consider the effect of instructional time at the school level as this is where timetables are set, and schools do retain the ability to respond individually to students as the need arises.

To construct the Carroll model, Opportunity to Learn, or 'instructional time', is calculated by summing the amount of time taught per EBacc subject group, as described in equations 2 and 3. The same is possible for the total hours taught per student for all KS4 subjects. Then, proxies are generated for Aptitude (prior cohort performance at KS2), Perseverance (levels of attendance) and Quality of Instruction (Ofsted rating). Unfortunately, no proxy for Ability to Understand Instruction is available, although covariates for prior learning at KS2, the percentage of students

with special educational needs, and the percentage of students with English as an additional language (EAL) are included. To test for a quadratic functional form, hours per student are standardized and squared. Thus, the model for the average value-added performance of students in department i, at time t (VA_{it}) is:

$$VA_{it} = \delta_t + \gamma_{it} + \beta_1 TPS_{it} + \beta_2 TPS_{it}^2 + X_{it} + u_i + \varepsilon_{it} \qquad (1)$$

where:

- δ_t are year FE for census years 2010–14
- γ_{it} is a dummy variable for Ofsted Quality of Teaching grade
- TPS_{it} is the time variable (or TPC_{it} – see equation 3 below), for department i in year t
- TPS_{it}^2 is the corresponding quadratic term
- X_{it} are time-varying department-level covariates for department i in year t
- u_i are time-invariant unobserved effects for department i (these will drop out when FE are applied)
- ε_{it} are the unobserved idiosyncratic errors for department i in year t.

The following seven covariates are used. As the English national curriculum makes English, mathematics and science compulsory, we expect cohort-level variables to apply at the department level (for example, KS2 average point score (APS)). Some caution exists in science and the humanities, as not all students are entered, although I attempt to control for this by adding covariate seven to the model:

1. student poverty levels (FSM)
2. proportion of students with EAL
3. proportion of students with special educational needs (SEN)
4. school size
5. KS2 APS – the average KS2 score for each cohort (as a proxy for aptitude)
6. absence levels (as a proxy for perseverance)
7. the proportion of Year 11 students entered for this EBacc pillar.

Time variables

Prior studies of allocated instruction time measure the number of classes and minutes per class, as is permitted by cross-sectional PISA data (Lavy, 2015; Rivkin and Schiman, 2015). The SWC does not directly contain these data, although it is approximated by equation 3. Instead the census holds teachers' timetable data, showing the total number of hours taught in a subject, to a particular year group. To the best of my knowledge, this is a novel estimate of allocated instruction time. Ten such variables are constructed (see Table 1). The first four measure Time per Student (TPS), which calculates the amount of KS4 time scheduled per EBacc subject dividing by the number of students present at the end of KS4 (equation 2). This estimates students' time over the two years of KS4 spent in that subject:

$$TPS_{it} = \frac{\Sigma KS4\ Hours_{it}}{Number\ of\ students\ at\ end\ of\ KS4} \qquad (2)$$

For department i in census year t.

The next four variables measure Time per Class (TPC), as given by equation 3. This estimates the Year 11 time allocated per Year 11 class for the four subject areas (English, science, mathematics and humanities). Year 11 is chosen as it is the final year

Table 1: Subject hours (2010–14)

Per student hours (all subjects)	Mean hours	SD
KS4 (all subjects)[#]	2.26	0.33
Yr11 (all subjects)	1.16	0.18
TPS (per student)		
KS4 English[#]	0.32	0.07
KS4 science[†,#]	0.40	0.09
KS4 maths[#]	0.31	0.07
KS4 humanities[#]	0.17	0.05
TPC (per Yr11 class)		
Yr11 English per group (28)	4.29	1.18
Yr11 science per group (28)[†]	5.25	1.36
Yr11 maths per group (28)	4.11	1.14
Yr11 humanities per group (28)	2.09	0.80

Source: SWC 2010–14, DQ level 20 per cent excluding outliers
[#]KS4 hours per student reflect two years of education
[†]Science hours exclude applied science (dual award)

of lower secondary school (International Standard Classification of Education Level 2), and it may help schools to visualize the possible impact of an extra hour of time, per class, in Year 11. This assumes an arbitrary class size of 28 rounded up to produce an integer number of classes.

$$TPC_{it} = \frac{\sum Year\ 11 Hours_{it}}{Roundup\left(\dfrac{Number\ of\ students\ at\ end\ of\ KS4}{28}\right)} \quad (3)$$

For department *i* in census year *t*.

Panel construction and estimation method

While experimental data are considered the gold standard for identifying causal effects, quasi-experimental methods are considered 'the next best thing' in the absence of such data (Wooldridge, 2010). This includes FE models as proposed in this study. Thus, analysis is conducted in Stata version 16 using the reg and xtreg commands. A wide panel is constructed with one row per school per year. This contains variables for 'KS4 hours per student' and 'hours per Year 11 class' in English, science, mathematics and humanities. So too, it contains value-added results for each subject and the school-level performance measures ptac5 and ptac5em. Finally, it contains school-level covariates which are also applicable at department level. In total, 2,815 schools are included, although the panel is unbalanced, with only 1,904 schools having data for four or five years; this is attributed to the opening and closing of schools, and years where schools did not meet data quality requirements. Accordingly, 22 different versions of the dependent variable are analysed in total. These allow us to estimate the effect of time on value-added performance in each subject area, for distinct sub-groups, including students with low/average/high prior attainment at KS2, and for students with FSM. This acts as a robustness check.

To answer research question one, I first calculate a baseline using OLS regression to estimate the likely impact on value-added in subjects separately. For

a more robust estimation, random effects generalized least squares (GLS) regression is also performed. Estimating subjects separately eliminates the need to control for clustering by school and avoids concerns of assessment comparability across subjects (Coe, 2008). Although this assumes that departments are independent, it does allow for the possibility that the effect, and functional form of time may vary by subject, and offers straightforward interpretation. It also allows the application of school-by-subject effects where the variation analysed is the variation in a subject department over time – for example, where departments act as their own controls. This minimizes 'omitted variable bias', as all time-invariant factors will drop out of the model. To answer research question two, the above is repeated but with new dependent variables, which show value-added attainment by sub-group. In all cases, robust standard errors are calculated. The use of value-added measures and the application of year FE are also intended to address concerns on possible 'grade inflation'.

Threats to internal and external validity

To ensure robustness, I first investigate issues of data quality, missing data and the potential for measurement errors. I then consider the sensitivity of findings to changes in the specification of variables and estimation method. Finally, I consider potential biases due to sorting.

There is good reason to have confidence in the SWC data, as DfE statistics are judged to be 'of the highest quality' (OSR, 2017). Starting in 2010, data are automatically collected from school databases in the annual November return and are validated both before and after collection. This suffers none of the typical issues associated with self-report surveys found in previous studies. Data are also collected for between 75 and 79 per cent of all teachers in English maintained secondary schools, making this a particularly large sample.

The missing data above are partly explained by the 14 per cent of schools who submitted no curriculum data between 2010 and 2014. A further 30 per cent of schools submitted data for some, but not all, census years. When comparing schools with complete data against schools without, there are no statistically significant differences observed by Ofsted rating, the proportion of students for whom English is an additional language and the average salary paid to staff (all comparisons are made in the first year schools are observed in matched records). However, significant differences showed that schools with missing data have a slightly higher percentage of students on FSM (+2.73 per cent, SD=12.74), percentage of absences (+0.12 per cent, SD=1.54), a slightly lower KS2 APS baseline (–0.25, SD=1.68), percentage of students gaining five 'good' GCSE passes (–0.82 per cent, SD=12.13) and lower enrolments (–36.27, SD=365.38). Perhaps these differences are partly explained by the observation that schools with missing data are twice as likely to be sponsor-led academies, where academization can be an indicator of change, with associated disruptions in administrative systems. At a practical level, however, these differences appear small and unlikely to bias the analysis proposed in this article. A third and final source of missing data is possible. This arises when some schools submit returns containing teachers both with and without timetable data. This may reflect a growing number of 'non-teaching' teachers, such as deputy heads or pastoral staff. As the actual number of such staff is unknown, I analyse schools with up to 10 per cent, 20 per cent and 30 per cent of such staff, but find little or no statistically significant pattern in terms of the average hours per student, nor in terms of school-level and subject-level results. In this study, schools with more than 20 per cent 'non-teaching' staff are excluded for reasons of plausibility (I call this the 20 per cent DQ or data quality level).

Next, I consider the sensitivity of results due to the choice of dependent and independent variable. Across all models, changes in the specification of dependent variable do not show any unanticipated and pronounced variation. Results are also robust to changes in the main explanatory variable when comparing overall hours across all subjects (see Table 4) and hours in each subject; while the effect sizes at school level are smaller than that seen at subject level, they are consistent with the overall picture.

The FE models used make certain exogeneity assumptions to ensure consistent estimators – for example, independent variables must be independent of the dependent variable. Put another way, we ask if the level of allocated instructional time (or any covariate) is dependent on the value-added obtained in previous years (for sequential exogeneity) or in any year (strict exogeneity). Consideration of the independent variables does suggest that the condition for strict exogeneity is likely to be met. One possible, but unlikely, violation may arise if a department underperformed one year and was then allocated additional instructional time the following year. This is unlikely, however, due to the timing of the timetabling process, which finalizes schedules before the publication of value-added results. Traditional OLS regression estimation also assumes an absence of heteroskedasticity and serial correlation. However, with large N and fixed T, the use of robust estimators allows this assumption to be relaxed (Wooldridge, 2010).

Following Lavy's (2015) approach, I also compare estimation methods and similarly see an upwards bias to OLS estimators (available upon request). This confirms the use of FE estimation.

Finally, it is crucial to consider the possibility of student sorting into schools. Were students to be sorted into schools based on subject-specific characteristics, this may result in biased estimates. For example, should students with an aptitude in modern foreign languages sort into schools with relatively more instruction time for languages, this would produce an upward bias. However, given a gap of five years between secondary school enrolment and KS4 results, this is highly unlikely. To test for the presence of a sorting effect, three strategies are enacted. First, average performance across all subjects and departments is assessed (see Table 4). The second strategy presumes that sorting effects will be stronger in London (via the logic that the concentration of schools makes moving schools easier) and thus looks for, but does not find, an interaction between location and hours per student (not shown). The final strategy compares the effect size of time in Year 9 as compared to Year 10, at which point 'sorting' might be expected to occur and lead to an increased effect size in Year 10 (available upon request). In all cases, no evidence for sorting was observed.

Before examining results, it is appropriate to evaluate whether any findings may support causal claims. In this study, the primary strategy used to identify a likely causal effect is the use of FE, where schools act as their own controls, thus eliminating any bias due to time-invariant changes. To control for time-varying exogeneity, a set of 'rich controls' are applied. Also, the use of administrative data either removes or reduces the chances of various typical forms of error and bias. Further support for causality comes from the checks described above. Finally, the effect sizes observed are also broadly consistent with findings from other jurisdictions. In this context, tentative claims for causality are suggested, and I adopt the phrase 'likely causality'.

Results

Results for OLS estimation yielded slightly larger effect sizes compared to longitudinal models, as expected. For each of English, mathematics and science, the OLS

estimators showed a statistically significant non-linear effect, although the sign of the quadratic term was unexpectedly positive. For humanities, a negative quadratic effect was observed. Even so, these estimates are generally smaller than was suggested by Scheerens et al. (2013), or indeed John Hattie (2015) and his update to time on task; this may be expected when dealing with administrative data for nationally standardized examinations (Kraft, 2018).

As more robust models are estimated, first using random effects, then department-level FE, the quadratic coefficients become non-significant and the overall effect size declines (available upon request); this suggests an upwards bias in OLS estimators when used on cross-sectional data, as indicated by Scheerens et al. (2013). Comparing random versus FE, there does appear to be relative agreement between results for English and mathematics; however, FE estimates in science and the humanities appear smaller, perhaps reflecting exogeneity that is better controlled by the application of FE where departments act as their own controls. This is not unexpected, as FE estimators are generally held to be more robust (Wooldridge, 2010). Figure 1 illustrates these results, showing the FE regression coefficient for KS4 hours per student, with 95 per cent confidence intervals and the effect size of time for each sub-group (given as a black bar). This demonstrates a consistent picture of a small effect size (between 0 and 0.1SD) irrespective of subject or student sub-group.

These regression results are listed in Table 2, which lists the effects of additional EBacc subject time, per student, on subject value-added for each cohort and
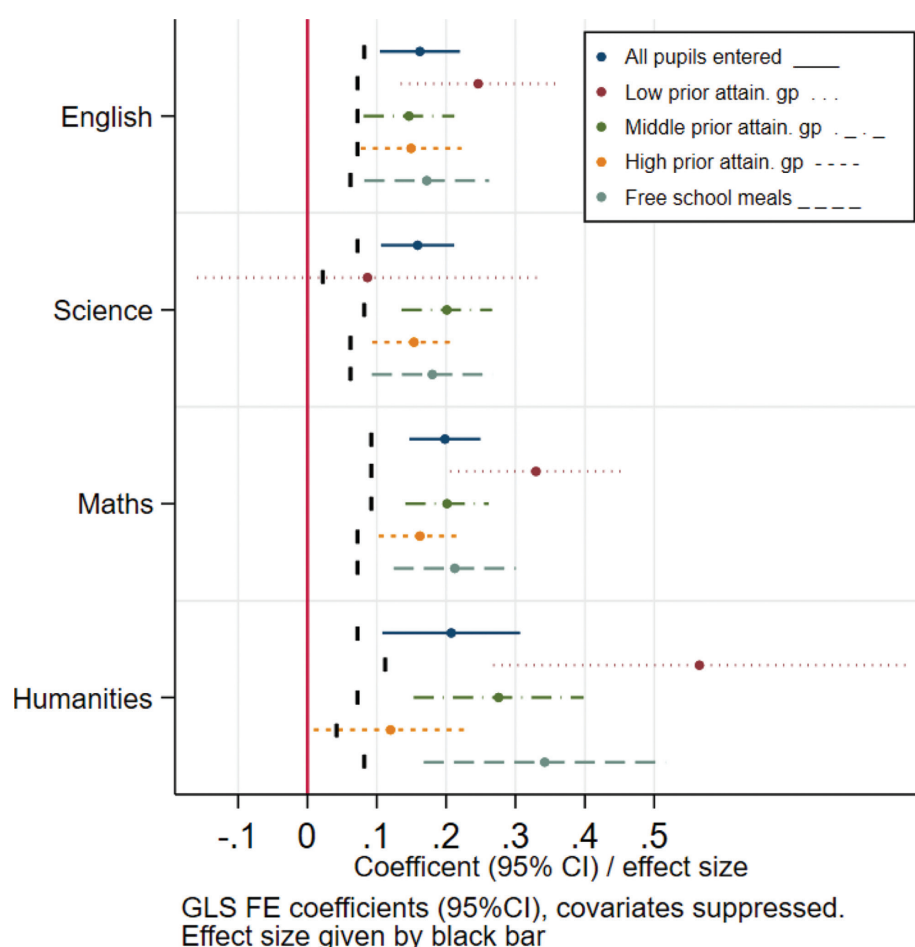

Figure 1: Regression coefficients and effect size for KS4 hours/student (by subject and sub-group) (Source: SWC, 2010–14, and matched data)

sub-groups based upon prior attainment and FSM status. The models control for the effects of percentage absence, students with FSM, SEN or EAL and cohort prior attainment, at the group level. The models also control for schools' Ofsted rating, year effects and time-invariant school-subject effects. Estimation results show a highly significant linear relationship between hours per student and subject value-added; however, little or no evidence exists for a non-linear relationship after the application of FE. The linear effect size of instructional time is small and in line with prior 'high quality' findings. Of the four subject areas examined, the effect size for mathematics was the highest (0.09SD for all students) and humanities (0.11SD for students with low prior attainment); effect sizes for English and science were somewhat smaller (English 0.08SD, science 0.07SD for all students). When examining student sub-groups, effect sizes were slightly smaller for students with stronger academic foundations at KS2. The effect size for groups of students with FSM appears lower than the effect size for the entire cohort.

To contextualize this effect size, three comparisons are made, as demonstrated by Figure 2. First, looking at the 'FSM gap', defined as the difference in value-added attainment between students with FSM and students with middle prior achievement, the effect sizes in Table 2 indicate that a one SD increase in KS4 hours of English may close the FSM gap by 6.5 per cent. In science, mathematics and the humanities, the gap is reduced by 8 per cent, 8.2 per cent and 9.3 per cent, respectively. This time investment equates, respectively, to an extra 59, 74, 57 and 44 minutes per week, per class of 28, for two years. Second, it is possible to ask whether additional time could narrow the value-added gap between outstanding and good schools, as judged by Ofsted. Here, a 1SD increase in KS4 time is equivalent to 13 per cent, 10 per cent, 14 per cent and 10 per cent, respectively, of the Ofsted gap. Finally, I compare the difference in value-added attainment between schools with above-average and below-average KS2 average points. Against this measure, a 1SD increase in time in KS4 reduces the attainment gap by 21 per cent, 13 per cent, 22 per cent and 9 per cent, respectively.

To assist in the practical implementation of these data for schools and to facilitate comparison with Lavy (2015) and Rivkin and Schiman (2015), I also estimate the effect of increasing instruction time by one hour per group of 28 students in Year 11. Using department and year FE, the estimation results are shown in Table 3 and are very close to results seen in Table 2, lending further support to the robustness of results. Here, the effect of one additional hour of instructional time, per Year 11 class, leads to an increase of value-added of 0.12, 0.09, 0.18, and 0.43 for English, science, mathematics and humanities, respectively. This equates to an effect of 0.06, 0.04, 0.08 and 0.14SD per hour. At a practical level, this seems small, particularly when considering the cost of such time.

It is also practically helpful to consider whether time in Year 9, 10 or 11 alone has likely impact. Thus, the EBacc subject hours per student are calculated for each
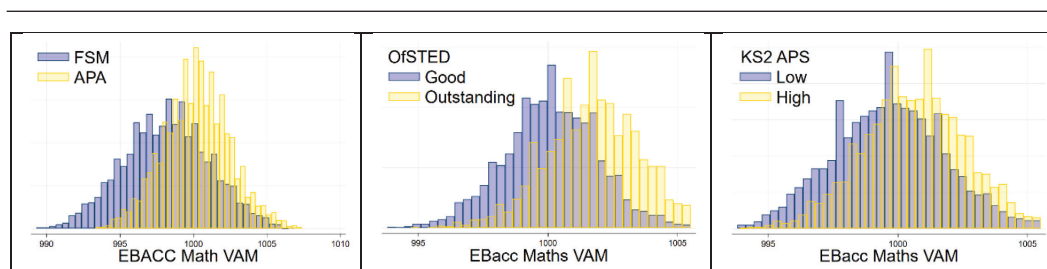

Figure 2: Gap between FSM and APA, Ofsted Good and Outstanding schools, and low and high KS2 APS (Source: SWC, 2010–14, and matched Ofsted data)

Table 2: Regression results for value added by sub-group and KS4 hours (2010–14)

English value-added	All students	Low prior attainment	Avg. prior attainment	High prior attainment	Students with FSM
KS4 hrs/pp (std)	0.16(.03)***	0.25(.06)***	0.15(.03)***	0.15(.04)***	0.17(.05)***
Hrs squared	0.01(.01)	0.00(.03)	–0.01(.02)	0.00(.02)	0.02(.02)
Effect size (SD)	0.08	0.07	0.07	.07	0.06
Number of schools	2,698	2,543	2,660	2,691	2,653
Science value-added					
KS4 hrs/pp (std)	0.16(.03)***	0.09(.13)	0.20(.03)***	0.15(.03)***	0.18(.04)***
Hrs squared	0.00(.02)	–0.12(.07)*	0.01(.02)	0.00(.02)	0.00(.02)
Effect size (SD)	0.07	0.02	0.08	0.06	0.06
Number of schools	2,681	1,619	2,641	2,676	2,595
Mathematics value-added					
KS4 hrs/pp (std)	0.20(.03)***	0.33(.06)***	0.20(.03)***	0.16(.03)***	0.21(.05)***
Hrs squared	0.00(.01)	0.02(.03)	–0.01(.02)	–0.01(.02)	0.02(.02)
Effect size (SD)	0.09	0.09	.09	0.07	0.07
Number of schools	2,684	2,532	2,648	2,692	2,658
Humanities value-added					
KS4 hrs/pp (std)	0.21(.05)***	0.57(.15)***	0.28(.06)***	0.12(.06)**	0.34(.09)***
Hrs squared	0.03(.03)	0.01(.08)	0.02(.04)	–0.01(.03)	–0.02(.05)
Effect size (SD)	0.07	0.11	0.07	0.04	0.08
Number of schools	2,693	2,088	2,652	2,674	2,579

Source: SWC 2010–14 (Curriculum table), School Census, Performance Tables and Ofsted results
All estimates calculated with school-subject FE, and year and Ofsted dummies
Excluded: Schools outside 1st and 99th percentile, > DQ level 20%, roll<300, Yr11<25 and leverage flag.
Significant levels given as: *0.05; **0.01; ***0.001 level. SE are given in parentheses.

Table 3: Effect of one additional hour per Year 11 class, by sub-group (2010–14)

Effect (SD)	All	Low prior achv.	Avg. prior achv.	High prior achv.	Students with FSM
English	0.06	0.05	0.05	0.06	0.04
Science	0.04	ns	0.04	0.04	0.04
Mathematics	0.08	0.08	0.07	0.07	0.07
Humanities	0.14	0.15	0.13	0.09	0.14[†]

Notes and source as per Table 2
[†]Approximation based on the non-linear effect
All results significant at the 0.001 level, unless indicated by 'ns'

of Year 9, Year 10 and Year 11. These are then considered in two ways. First, all three measures are entered together into the regression equation. Second, they are entered separately. In the first instance, entering all three time measures simultaneously yields a significant effect for Year 11 time, but not for Year 9 or Year 10 time (not shown). This result is consistent across subject and student sub-groups. When considered

Table 4: Effect size (SD) of KS4 instruction time and attainment at school level (2010–14)

	ptac5[†]	ptac5em	ptac5	ptac5em
Effect size of KS4 hours per student (SD)	0.040***	0.020***	0.025*	0.018***
GLS-random effects	Y	Y	N	N
School FE	N	N	Y	Y
Number of schools	2,713	2,711	2,713	2,711

Notes and source as per Table 2
Significant levels given as: *0.05; **0.01; ***0.001 level
[†]pt of students achv 5+ A*–C or equiv (ptac5em = including English and mathematics)

separately, each time measure typically has a significant effect size (available upon request), and while these are broadly similar, the estimates for Year 11 do appear slightly larger. The exception to this is for humanities subjects, where time in Years 9 and 10 is non-significant, while time in Year 11 has a considerably larger effect.

Finally, as a pre-planned robustness check, aggregate hours across KS4 subjects are calculated and regressed on school-level performance indicators ptac5 and ptac5em, first using GLS-random effects and then FE estimators where schools act as their own controls. As shown in Table 4, the observed FE effect sizes (0.025 and 0.018SD) appear smaller than seen in Table 2; this is likely due to a less direct link between hours aggregated across all KS4 departments and cohort performance measures. Nevertheless, this finding is similar to that seen at the subject level, which supports the robustness of results in Tables 2 and 3.

Discussion

This article makes a distinct and new contribution to the literature by examining the association between additional instruction time and English secondary schools' KS4 value-added performance in national GCSE examinations. Examining English, science, mathematics and the humanities separately, using longitudinal analysis with school-by-subject FE, the possible effect of such additional time is small and varies across subject (0.08, 0.07, 0.09 and 0.07SD, respectively). Based on additional time in both Years 10 and 11 (KS4), mathematics appears to benefit most and the humanities least. However, a slightly different picture appears if analysing additional time in Year 11 alone, which seems to have a higher likely impact than additional time in Year 10, particularly for the humanities (effect size = 0.12SD), perhaps reflecting proximity to GCSE examinations.

School reformers and policymakers have praised 'additional time' as an essential tool in the school improvement arsenal. So too, it has been feted as part of the success of highlighted American charter schools and English academies (Gove, 2012). By extending their school days, weeks and years, these schools offer between 20 and 100 per cent additional instruction time (DfE, 2014; Dobbie and Fryer Jr, 2013). However, in these studies of successful innovation, extra time has been but one of many related interventions to address disadvantage and poverty. In contrast, the results of this article offer pause for thought for English secondary schools considering extending their school day, or reallocating curriculum time, as a means of raising attainment.

While small returns to extra time seem counter-intuitive, there are multiple possible explanations. Before exploring those, it is crucial to say that this result does not indicate that the absence of any instructional time would still result in GCSE success. Instead, considering the nature of this data set, I only comment upon the likely effect of *additional* time. As both theory and prior evidence suggest diminishing returns to

time (Scheerens et al., 2013), it is plausible that such diminishing returns could explain what seems like a small effect. To test this possibility, a quadratic term was included in the regression models; however, this did not yield consistently statistically significant results. While this lack of significance poses a challenge to this explanation, my findings are still conceivable if the underlying curve is mainly linear through the range of hours per subject/student typically covered in England's schools. This is a question that must be left to further research.

Other explanations may link to variation in pedagogy and student motivation, although they do seem highly unlikely in such a large data set, spanning five years. Even though I have attempted to control for quality of teaching through the inclusion of Ofsted ratings, this school-level measure is somewhat imprecise and not always very current. Therefore, it is possible that teachers with less time employ different pedagogical tools that have larger effect sizes. Such an effect would result in an underestimation of the effect size of instructional time, *ceteris paribus*. However, were this the case, it would suggest a benefit to giving teachers additional non-contact time to prepare/learn such techniques.

There may also be a benefit to students having less contact time, if this enabled more productive use of non-contact time. This benefit was demonstrated in a three-year longitudinal study of a 'learning to learn' programme for one KS3 cohort in an English secondary school. In this experimental study, students received some four hundred periods fewer in their KS3 curriculum in order to study 'learning to learn'; nevertheless, their end of KS3 results were still markedly higher than the control group (Mannion and Mercer, 2016). Conversely, it is also plausible that students with higher contact hours feel overburdened and pressured, leading to underperformance. This is consistent with indicators of low life satisfaction among England's secondary students, as compared with their international peers (OECD, 2017).

As schools adjust to COVID-19, this study has clear implications which offer some hope. If schools are not able to manage 'normal' levels of teaching, small reductions in time are likely to have small impact on GCSE attainment at a cohort level. If schools must offer a reduced timetable to ensure student safety, it is conceivable that forgone curriculum time may be reapplied in ways to mitigate any likely impact; such methods are discussed in the Education Endowment Foundation's (EEF, n.d.a) influential *Teaching and Learning Toolkit*. Applying the guidelines used by the EEF (n.d.b), the effect of a 1SD reduction in instruction time is likely to equate to between one and two months of progress, depending on student background and subject. However, the EEF toolkit highlights the considerably larger benefits that can arise from the use of feedback (eight months of progress), meta-cognition and self-regulation (seven months), homework (five months), collaborative learning (five months), peer tutoring (five months) and parent engagement (three months of progress). Thus, with the application of a refined curriculum and pastoral tools, this use of time could see overall benefits to attainment. A similar conclusion was offered by John Hattie (2020) when considering the effects of the 2011 earthquake in Christchurch, New Zealand.

As previously indicated, my findings are cautious and limited by the weakness of this study. Not being based on experimental data, I only report 'likely effects'. The use of national administrative data also comes at a cost, as the data used are school-level data, rather than student-level data, making the ecological fallacy a possibility. This fallacy appears unlikely however, as the pattern seen across both subjects and student sub-groups is consistent. At a practical level, it is possible that exceptions to the rule will exist for some students, but this point seems uncontentious. Notwithstanding, analysis with student-level data would allow for a more robust linkage between time

and attainment, and greater assessment of student differences. Also, these results do not necessarily apply to primary-aged children and nor do they consider important non-examination-based outcomes.

Time in schools has long been an area of research interest, perhaps helped by the controversial 1966 'Coleman report'. Notwithstanding, David Berliner (1991: 3) commented in 1991 that some 'scholars find the concept of instructional time to be intellectually unexciting … commonsensical … trivial … [with] findings that have the status of truisms (e.g., students who spend more time studying learn more)'. At one level, as Berliner (1991) illustrates, to study the use of time in the classroom is to study effective teachers at work. At a different level, to question the allocation of time is to ask questions about the effectiveness of resource allocation in schools. Decisions about time allocation are fundamental reflections of priorities for curriculum, staff training and student welfare – time given to one focus is time omitted from another. Such issues do not seem trivial, particularly in England, where schools have the freedom to allocate time as they see fit, whether it be throughout the school day, week or year. In 2021, this is highly pertinent as schools respond to COVID-19. Echoing John Carroll's (1989) sentiment, the results of this analysis suggest that schools may find it more productive to consider carefully the range and quality of activities provided, as opposed to the quantity. In the years to come, as schools 'build back better', 'more time for all' is an unlikely answer to calls for improving results, and 'narrowing the gap', and a re-evaluation of the use of time in schools may allow for previously unrealized benefits and gains.

Acknowledgements

Notes on the contributor

Vaughan Connolly is a PhD candidate in the Faculty of Education, University of Cambridge, UK.

References

Berliner, D. (1991) 'What's all the fuss about instructional time?'. In M. Ben-Peretz and R. Bromme (eds), *The Nature of Time in Schools: Theoretical concepts, practitioner perceptions*. New York: Teachers College Press, 3–35.

Berliner, D. and Biddle, B. (1995) *The Manufactured Crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.

Bietenbeck, J. and Collins, M. (2020) *New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments* (Working Paper 2020: 18). Department of Economics, Lund University. Accessed 22 April 2021. http://lup.lub.lu.se/record/fef7da41-a969-4b23-9158-e074a826c149.

Bingley, P., Heinesen, E., Krassel, K.F. and Kristensen, N. (2018) *The Timing of Instruction Time: Accumulated hours, timing and pupil achievement* (No. 11807; IZA Discussion Paper Series). IZA Institute of Labor Economics. Accessed 22 April 2021. http://ftp.iza.org/dp11807.pdf.

Carroll, J.B. (1989) 'The Carroll model: A 25-year retrospective and prospective view'. *Educational Researcher*, 18 (1), 26–31. https://doi.org/10.2307/1176007.

Coe, R. (2008) 'Comparability of GCSE examinations in different subjects: An application of the Rasch model'. *Oxford Review of Education*, 34 (5), 609–36. https://doi.org/10.1080/03054980801970312.

DfE (Department for Education) (n.d.) 'Find and compare schools in England: Download data'. Accessed 27 July 2019. www.compare-school-performance.service.gov.uk/download-data.

DfE (Department for Education) (2014) *ARK Schools. In-depth sponsor profile*. Accessed 22 April 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/313463/ARK_Schools_Case_Study_FINAL_APPROVED.pdf.

DfE (Department for Education) (2015) *Key Stage 2 to Key Stage 4 Value Added Measures: A technical guide for local authorities, maintained schools, academies and free schools: January 2015*. Accessed 22 April 2021. https://dera.ioe.ac.uk/25969/.

DfE (Department for Education) (2016) *Revised GCSE and Equivalent Results in England, 2014 to 2015: Quality and methodology information: SFR 01/2016*. Accessed 22 April 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/493295/SFR01_2016_QualityandMethodology.pdf.

Dobbie, W. and Fryer Jr, R.G. (2013) 'Getting beneath the veil of effective schools: Evidence from New York City'. *American Economic Journal: Applied Economics*, 5 (4), 28–60. https://doi.org/10.1257/app.5.4.28.

EEF (Education Endowment Foundation) (n.d.a) *Teaching and Learning Toolkit*. Accessed 11 August 2020. https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit.

EEF (Education Endowment Foundation) (n.d.b) *The EEF's Months of Additional Progress Measure*. Accessed 18 July 2020. https://educationendowmentfoundation.org.uk/help/projects/the-eefs-months-progress-measure/.

EEF (Education Endowment Foundation) (2019) *Extending School Time*. EEF Toolkit. Accessed 22 April 2021. https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/extending-school-time.

Fraser, B.J., Walberg, H.J., Welch, W.W. and Hattie, J.A. (1987) 'Syntheses of educational productivity research'. *International Journal of Educational Research*, 11 (2), 147–252. https://doi.org/10.1016/0883-0355(87)90035-8.

Gove, M. (2012) 'Michael Gove speech on academies'. 4 January. Accessed 22 April 2021. www.gov.uk/government/speeches/michael-gove-speech-on-academies.

Hanushek, E.A. (2015) 'Time in education: Introduction'. *The Economic Journal*, 125 (588), F394–F396. https://doi.org/10/ghctqt.

Hattie, J. (2009) *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement.* Abingdon: Routledge.

Hattie, J. (2015) 'The applicability of Visible Learning to higher education'. *Scholarship of Teaching and Learning in Psychology*, 1 (1), 79–91. https://doi.org/10/gcx4mj.

Hattie, J. (2020) 'Visible Learning effect sizes: When schools are closed: What matters and what does not'. *Corwin Connect*, 14 April. Accessed 22 April 2021. https://corwin-connect.com/2020/04/visible-learning-effect-sizes-when-schools-are-closed-what-matters-and-what-does-not/.

Karweit, N. (1984) *Time-on-Task Reconsidered: Synthesis of research on time and learning*. Accessed 22 April 2021. www.ascd.org/ASCD/pdf/journals/ed_lead/el_198405_karweit.pdf.

Kraft, M.A. (2018) *Interpreting Effect Sizes* (Brown University Working Paper). Accessed 22 April 2021. https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf.

Lavy, V. (2015) 'Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries'. *The Economic Journal*, 125 (588), F397–F424. https://doi.org/10/gbf7z2.

Mannion, J. and Mercer, N. (2016) 'Learning to learn: Improving attainment, closing the gap at Key Stage 3'. *The Curriculum Journal*, 27 (2), 246–71. https://doi.org/10.1080/09585176.2015.1137778.

OECD (Organisation for Economic Co-operation and Development) (2017) *Are Students Happy? PISA 2015 results: Students' well-being*. PISA in Focus, No. 71. https://doi.org/10.1787/3512d7ae-en.

OSR (Office for Statistics Regulation) (2017) *Assessment of Compliance With the Code of Practice for Official Statistics: Statistics for England on schools, pupils and their characteristics, and on absence and exclusions* (Assessment Report No. 332). Accessed 25 April 2021. https://osr.statisticsauthority.gov.uk/wp-content/uploads/2017/02/Assessment-Report-332-Statistics-for-England-on-Schools-Pupils-and-their-Characteristics-and-on-Absence-and-Exclusions.pdf.

Patall, E.A., Cooper, H. and Allen, A.B. (2010) 'Extending the school day or school year: A systematic review of research (1985–2009)'. *Review of Educational Research*, 80 (3), 401–36. https://doi.org/10.3102/0034654310377086.

Rivkin, S.G. and Schiman, J.C. (2015) 'Instruction time, classroom quality, and academic achievement'. *The Economic Journal*, 125 (588), F425–F448. https://doi.org/10/gbf7z3.

Scheerens, J., Luyten, J.W., Steen, R. and de Thouars, Y.C.H. (2007) *Review and Meta-Analyses of School and Teaching Effectiveness.* Enschede: Universiteit Twente.

Scheerens, J., Hendriks, M., Luyten, H., Sleegers, P. and Glas, C. (2013) *Productive Time in Education: A review of the effectiveness of teaching time at school, homework and extended time outside school hours.* Enschede: Universiteit Twente. Accessed 22 April 2021. http://doc.utwente.nl/86371/1/Productive_time_in_education.pdf.

Van Damme, D. (2014) 'Is more time spent in the classroom helpful for learning?' *OECD Education Today.* 13 May. Accessed 22 April 2021. http://oecdeducationtoday.blogspot.it/2014/05/is-more-time-spent-in-classroom-elpful.html.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.