The afternoon effect: differential impacts on student performance in maths and history

Velichka Dimitrova

Abstract

Understanding the determinants of cognitive achievement and improving the efficiency of the educational system requires knowledge the time-of-day effects on the learning process. As productivity may vary during different times of the day depending on academic task at hand, there may be different optimal times to schedule a maths, history or language class, respectively. Using a decade-long panel dataset of academic scores in different subjects from a humanities high school, I estimate value-added and fixed effects specifications of a dynamic educational production function. Exogenous variation from random allocation to morning and afternoon school start times enables the identification of an afternoon effect. The findings indicate significantly lower maths scores during afternoon classes by 0.082 (0.018) and higher test scores for history classes by 0.069 (0.029) standard deviations, respectively. Using a natural experiment of a transitioning from a double-shift to a morning-only school schedule, I estimate difference-in-difference model of regime change. In a quantile regression, I investigate the distinct impacts of the covariates at different segments of the grade distribution and find that students with the lowest grades stand to lose most by having classes scheduled in the afternoon. These results point to a cost-effective way to achieve better academic performance with more effective organisation of school inputs accounting for the time-of-day effects.

Key words*: time-of-day effects, optimal schedules, school start times, afternoon effect, educational quality, educational production*

I. Introduction

Increasing productivity and efficiency in the educational system requires targeted application of resources and better understanding of the time-of-day effects on the learning process. A large body of research on educational production functions considers how educational resources, including both school and non-school inputs are transformed into academic outcomes[1]. An important simplifying assumption is no heterogeneity in the school input effects for an individual student. Emerging evidence from new studies[2] demonstrates that heterogeneous responses to school inputs complicates the identification of these effects. In the context of time-of-day effects, students may not be able to maximise their learning opportunities and would absorb the same educational inputs in a different way, depending on varying productivity for academic tasks during the day and the constraints the school schedules impose on them.

A growing number of recent studies have considered how the time-of-day may affect learning, but there is remaining ambiguity. While the school start times literature has highlighted that later start and even afternoon classes may have a positive impact on student performance[3], studies of double-shift schooling[4] point to lower academic achievement for later-in-the-day classes versus morning classes for some subjects and inconclusive evidence for others Lusher and Yasenov (2016); Pope (2016). Time-of-day effects have been subject to research in psychology and neuroscience: to determine how human physiological and cognitive performance varies at different times of the day and how it depends on the circadian rhythm, a person's chronobiology and the effects of the light-dark cycle[5]. It has been widely accepted now that sleep deprivation can significantly impair memory and the acquisition of skills (e.g. Cardinali, 2008).

The literature is in agreement that sufficient sleep is a necessary prerequisite for performing well in cognitive tasks. In education, later secondary school start times have gained momentum with a number of medical and educational institutions urging schools to abandon early school start times. At the same time, double-shift school schedules are also popular educational policies. Currently, over 45 countries use some system of alternation between morning and afternoon classes[6]. In an extensive review Bray (2008) argues that the supply of school places can be increased by alternating school places, making better use of school resources and making a significant contribution to the provisions of universal primary and secondary education. Less is known, however, on how substantial delays in start times can affect the quality of educational inputs, including the way teachers deliver and how students perceive learning in different times of the day.

This study compiles a decade-long dataset of a humanities high school, containing term GPAs for a range of subjects and absences. The high school operated a school schedule with alternation between morning and afternoon classes until a building extension enabled a transition to morning-only start times. The random assignment of the school shifts and the alternation meant that a student

[1] Identifying assumptions and summary of studies in Meghir and Rivkin (2011); Todd and Wolpin (2003).

[2] Choi et al. (2014); Ding and Lehrer (2014); Lazear (2001).

[3] Carrell et al. (2011); Dills and Hernandez-Julian (2008); Edwards (2012); Hinrichs (2011).

[4] A school system where two separate groups of students use the same school facilities, building and is taught by the same teachers but attend school during different times of the day. The first group attends classes from morning until mid-day and the second from mid-day until evening (Bray, 2008).

[5] Craig and Condon (1985); Goldstein et al. (2007); Mackenberg et al. (1974); Miller et al. (1992); Monk and Leng (1982); Song and Stough (2000).

[6] (Lusher and Yasenov, 2016) Additionally, growing pupil numbers and and a limited number of schools in the United Kingdom, also was a reason for the National Education Trust to call for schools to consider double-shift schooling (Blatchford, 2015).

had exposure to afternoon classes in one of the two academic terms each year. The time series of individual test scores enables the estimation of a dynamic educational production function with value-added specification of school inputs, including an autoregressive component for an individual subject and an overall average term GPA as a pre-determined covariate. The long time dimension per student allows for isolating the individual-specific heterogeneity (academic capacity, effort, chronotype). Then, the exogenous regime change to a different schedule system enables observing cohorts under both double-shift and morning-only start times and estimating the change in the schedules with a difference-in-difference specification.

Research in this area seeks to reveal whether there is a optimal time to schedule a maths class, a language class or a history class and whether having classes later during the day would have differential impacts on the academic achievement of a student in different subjects. As with the discussion of class size (Lazear, 2001), it could become optimal to sort students differently according to the differential time-of-the-day effects or organise school schedules with certain subjects earlier in the day than others. Re-arranging school schedules in a more optimal way does not require investment of additional resources and could be a cost-effective intervention leading to increased academic performance.

Section II sets out the background and the relevant literature including a discussion of the methodological issues in estimating the impacts of educational inputs and the studies of shift-work and productivity, school start times, sleep and academic achievement. Section III describes the dataset and the educational context. Section IV presents the methodology and identification strategy. Section V describes the findings and the robustness checks, Section VI presents a interpretation and discussion of the finding and finally Section VII concludes.

II. Background

Estimating time-of-day effects on cognitive achievement follows significant progress in the educational production and school quality literature. Growing body of research has accumulated evidence on the the impact of class size[7], teacher quality and value-added[8] and school quality more generally[9]. Beyond the traditional school inputs, recent research looks into what makes effective schools beside the traditional resource inputs. School characteristics, instruction length and school organisation explain a large part of the variation in test scores (Angrist et al., 2013; Dobbie and Fryer, 2013).

Identification in educational production and educational quality has established several better research design principles, which have become part of modelling academic achievement, including the value-added specifications and the use of long panels (or siblings and twins), which enable differencing of individual-specific heterogeneities. Outcomes acquired in one period tend to persist into future periods, which reflects the technology of skills as self-reinforcing, augmenting the skills produced in later stages (Cunha and Heckman, 2007). Given the historical school and family inputs into the educational production, model specifications have to account for learning as a cumulative process (Cunha et al., 2010; Meghir and Rivkin, 2011; Todd and Wolpin, 2003). Studying the causal impact of educational inputs - how variations in educational quality may determine different returns to education - is often impeded by endogenous choices and unobserved heterogeneities. Without random assignment, self-selection into preferred course times would challenge uncovering the true parameter

[7] Angrist and Lavy (1999); Krueger et al. (1999).
[8] Araujo et al. (2016); Chetty et al. (2014); Lavy (2009); Rothstein (2010).
[9] Card and Krueger (1992); Das et al. (2013); Deming et al. (2014).

of interest (Dills and Hernandez-Julian, 2008; Pope, 2016). Synchrony effects between chronobiology type[10] and capacity for cognitive achievement during a particular time of the day would further complicate identification of the time-of-the-day effects (Goldstein et al., 2007).

The research of time-of-day effects is not limited to educational production and follows progress in the study of worker productivity and the costs and benefits of different schedules. When to schedule a lesson, a process or a work task is essentially a choice of production technology. As productivity may vary during different times of the day, there can be a variation in costs associated with different schedules. The studies of time-of-day effects on achievement and productivity consider how shift-work patterns may alter outcomes for workers and findings have implications for optimal scheduling. Shift work involving unusual work patterns and working at night is related to compensating wage differentials, exacting wage premiums for onerous work conditions (Hwang et al., 1992; Kostiuk, 1990; Lanfranchi et al., 2002). This literature also finds a large degree of self-selection and a bias related to unobserved worker productivity as well as variations in labour supply given time-of-day effects (Camerer et al., 1997; Farber, 2015; Hwang et al., 1992). Time-of-day effects are not limited to shift working and are observed in the outcomes of other daytime jobs, including high-skilled professionals. Extraneous factors related to the time-of-day including hunger, fatigue and breaks can influence decisions of judges (Chen et al., 2016; Danziger et al., 2011). Similarly, daylight saving and reduction of average sleep time has been associated with different behaviours of financial market participants and negative stock returns (Kamstra et al., 2000)

Shift work, working at unusual times or sleep deprivation may impede productivity. Folkard and Tucker (2003) point to reduced productivity and safety during night shifts citing impaired health, shortened sleep and disturbed social life as some of the mechanisms driving their findings. Sleep deprivation can inhibit cognitive capacity and executive control, which are the two core components of bandwidth[11]. Sleep deprivation and fatigue has been also linked to decreased cognitive performance and alertness and greater likelihood of making errors (Ulmer et al., 2009). Exogenously imposed variation in sleep patterns through Daylight Saving Time (DST) or sunset in different time zones has provided opportunities of studying the impact of sleep reduction on productivity and labour market outcomes. Barnes and Wagner (2009) find that exogenous sleep reductions through the Daylight Saving Time changes increase the incidents of workplace injuries. Using sunset time as a source of exogenous variation, impacting on bed times, Gibson and Shrader (2015) find that workers who sleep less (instrumented by later sunset times), have lower productivity and lower wages.

There is an aspect specific to adolescents and time-of-day effects in the educational context. Course scheduling and start times have been a subject in the school quality literature with the underlying mechanisms explained by the accumulating evidence in psychology research. The sleep needs and biology of adolescents are different from those of children and adults. There is a well-documented adolescent sleep phase delay, related to a tendency for later times both in falling asleep and waking up (Carskadon et al., 1993, 1998). The mechanisms of the sleep phase delay relate to the later circadian rhythm timing of adolescents and the production of melatonin - the hormone regulating the sleep onset. Mature teens demonstrate delay in the timing of the melatonin production, which explains both why they have a preference for staying up late and why it is more difficult for them to wake up in

[10] Chronobiology type refers to the preference for morning or afternoon schedules based on human biology and different propensities to sleep during a 24-hour period.

[11] Bandwidth refers to "the brain's ability to perform the basic functions that underlie higher-order behavior and decision-making" (Schilbach et al., 2016).

the morning (Carskadon et al., 1998). As adolescents do not have decreased need for sleep relative to children but tend to stay up late, early school start times result in them accumulating substantial sleep deficits (Wolfson and Carskadon, 1998). Sleep deprivation, including daytime sleepiness and irregular sleep patterns amongst adolescents have been linked to lower academic performance (Pagel et al., 2007; Wolfson and Carskadon, 1998) with more recent studies exploiting randomized assignment of students to courses and instructors to show that later school start or afternoon class has a significant positive effect on student achievement (Carrell et al., 2011; Dills and Hernandez-Julian, 2008; Edwards, 2012). These finding have led to to wider adoption of later school start times.

Using the variation generated by different school start times and double-shift schooling, recent studies argue for a detrimental effect related to studying in the afternoon (Lusher and Yasenov, 2016; Pope, 2016) with stronger evidence for maths scores. Having a class in the first period, however, can be related to reduction in the grades of that subject (Cortes et al., 2012). While Bray (2008) argues that the supply of school places can be increased by alternating school shifts, thus making better use of school resources, the time-of-day effects remain under-studied. There are possibly effects working in opposite directions: while afternoon classes may allow adolescent students to get sufficient sleep according to their chronotype preference, lunch productivity dip and afternoon fatigue may eliminate any benefits. Depending on the academic task, morning or afternoon schedule may have a different impact on test scores contingent on the cognitive task at hand. Mackenberg et al. (1974) find that performance in repetitive, automatised or overlearnt tasks was better in the morning, while "perpetual-restructuring" tasks had exactly the reverse pattern with better performance in the afternoon. Evidence from shift-work patterns with repetitive and automatised tasks also points to lower productivity during night shifts (Folkard and Tucker, 2003). School subjects may require different cognitive tasks and therefore the afternoon classes may have differential consequences.

While the chronobiology and the productivity time-of-day effects can be the main channel for the impact of afternoon classes, there are other factors which may affect cognitive performance. Variation in schedules also impose restrictions on social interactions as social organisation is based on a common understanding of time: activities have to synchronised with other parties. Time plays a major role in social interactions (Doleac and Sanders, 2015; Hamermesh et al., 2008) and external cues can alter the requirements on one's schedules. The schedules themselves impose restrictions on how students manage the rest of their day: block schedules may have advantages relative to the school day through scheduling instruction time later during the day (Rice et al., 2002). Attending classes in the afternoon can also change pupils' opportunities to attend extra-curricular activities and would interfere with homework and free time activities, which may introduce another channel of impacting test scores (Gentzkow and Shapiro, 2008).

III. Data and Context

This analysis uses student-level term data from a humanities high school in Bulgaria. Five cohorts study in the high school at any one year, the youngest admitted after middle school. The school admits pupils after the seventh grade - when they are 14 years old - and allocates them to a profile group capped at 26 people with a main subject in either literature, history, German, English or English + Greek, which is the main subject they study[12]. There is higher demand for foreign languages as a main

[12] Studying a particular main subject means that students get more teaching and specialised lessons, e.g. if a class studies English as a main profile subject, they may study also a regular subject like biology also in English.

profile subject and this is reflected both in the higher entry GPA for the high school admission[13], as well as higher GPAs in all subjects and lower number of absences for students from foreign language profiles relative to students from the profiles of literature and history (see Table A2 and A3 in the Appendix).

The school operated a double-shift school schedule due to capacity constraints until 2012. Students alternated between morning classes and afternoon classes in the first term (September to January) and the second term (February to June) respectively. Through this alternation schedule, each student had afternoon classes in one of the school terms. A building extension was completed in 2011, allowing transitioning of all students to morning-only start times in the next school year. This regime change had consequences for the classes scheduling and the length of time spent in school, but not for other school inputs and it did not result in hiring more teachers or investing additional school resources.

The data spans nine years (2008-2016) where all data for all grades within a term was digitalised from paper records (Table 1). Personal files of all student graduates contain the term GPAs in each subject and their absences. The subjects literature, mathematics, the first foreign language and sports were studied in all years, regardless of the profile. The absences divide into two categories: excused and unexcused, where excused absences usually result from sickness absence, while unexcused absences accumulate through being late or skipping class [14]. Table A3 demonstrates that while there was no significant difference in the excused absence rate between morning and afternoon classes, afternoon classes were associated with a significantly higher rate of unexcused absence rates: 0.058 (0.018) standard deviations during 2008-2011. Both absence rates were also significantly larger before the regime change and were higher during the second term of the school year. A higher average GPA was associated with both significantly lower excused and unexcused absences.

The data is complemented with school records on the entry GPA of students for admission in the high school (nationally-held examinations) as well as the matriculation exams results, also held nationally, which qualify a student to receive a high school graduation diploma and apply for a place at university. A high school student always writes two final matriculation exams: an obligatory one in literature and one optional in a subject of their choice.

Table A1 presents summary statistics for the full sample (2008-2016) and for subsamples pre-reform (2008-2011) and post-reform (2012-2016) as well as the morning and afternoon subsamples pre-reform. The afternoon term GPA results are higher on average for all subjects (including overall GPA) except for mathematics, where the maths term GPA is higher for morning classes. This difference is only statistically significant for the subjects mathematics and history. All subjects term GPAs are higher before the school schedules change in 2012, including mathematics and are statistically significant. There are also significant differences between the absences with the pre-reform sample having a lower number of excused absences but a higher number of unexcused absences. There are no significant differences in the gender composition.

The curriculum and the way students are tested is set out in national guidelines, and while there is not substantial difference between subjects on how students are tested - there are written exams in all subjects - there are differences between mathematics and the other humanities subjects with respect to

[13] Formed as an average of the results from the nationally held examinations in literature and mathematics and the final GPAs from the seventh grade.

[14] One absence is equivalent to missing one class, or being late three times. For estimation and interpretation, the absences are also standardised with mean zero and variance one.

Table 1: *Scheme of student cohorts with digitalised personal student files. The dataset spans nine school years between 2007/08 - 2015/16. The school year starts in September and ends in June the following year. For simplicity, the table and the text refers to school year 2007/08 as year 2008. A student cohort enters high school in the eighth grade and graduates in the twelfth - the final year of secondary school. For instance, a student cohort, which entered high school in 2008/09, graduated in 2013 and had three years of exposure to the double-shift classes schedule. The school year is divided in two academic terms: the first beginning in September and finishing at the end of January and the second beginning in February and ending in June. Prior to 2012, the school operated a double-shift schedule with alternation between morning and afternoon classes for all students. In 2011 a building extension enabled transitioning to morning only start-times. Colour legend: "cool black" - darker shade - (2008-2011) for double-shift schedules, "sea green" - lighter shade: (2012-2016) for morning-only start times.*

Year	’04	’05	’06	’07	’08	’09	’10	’11	’12	’13	’14	’15	’16	’17	’18	’19	’20
	8	9	10	11	12												
		8	9	10	11	12											
			8	9	10	11	12										
				8	9	10	11	12									
					8	9	10	11	12								
						8	9	10	11	12							
							8	9	10	11	12						
								8	9	10	11	12					
									8	9	10	11	12				
										8	9	10	11	12			
											8	9	10	11	12		
												8	9	10	11	12	
													8	9	10	11	12

the learning process and dynamics. Mathematics is exclusively a problem-solving subject, where the learning dynamics involve either the teacher demonstrating solutions in front of the class or giving students problems to solve and giving solutions at the end of the class.

History is the subject which involves the most intensive use of narrative: the teacher relates the subject matter and students take notes. The interaction with the teacher takes the form of answering questions and being tested. The literature and the foreign language subjects involve studying vocabulary and the structure of language, but are further away from the learning dynamics of a narrative. On the spectrum of academic tasks the subject of mathematics is arguably the more repetitive/automatised while history is closer to perpetual-restructuring mental activity. Mathematics teachers have complained that scheduling classes later in the school day works against the effectiveness of students, as they less able to concentrate and perform worse when mathematics classes are scheduled later during the school day.

Before 2012, during the double-shift system, all classes were 40 minutes, the first shift starting at 7:20am and ending at 1:30pm and the second shift starting directly after that and ending at 7:20pm. After 2012, in the single-shift system, classes began at 7:45am and were 45 minutes long, ending at 2:30pm or 3:15pm (if there were eight classes rather than the average of seven). The scheduling of the classes during the school day e.g. with morning-only start times can introduce additional variation: some classes may be very early in the day while others can be after lunch. However, this variation is balanced by having several classes a week where some classes could be an early morning start, while others can be later during the morning, yet on average these classes are coded as "morning". The same definition is valid for the afternoon classes, where some can be in the early afternoon, while others in the end of the day, yet there is balance on average as students would have several classes of the same subject in the same week at different times of the afternoon.

IV. Identification Strategy

The random assignment of morning and afternoon classes[15] and the variation generated within a student over time, enables identification of the time-of-the-day effects on cognitive achievement using a within estimation in a dynamic panel data setting with a value-added specification to account for learning as a cumulative process. The natural experiment context created by the transitioning to a new system with morning start times gives the opportunity to use difference-in-difference model of regime change. Using quantile regression I also estimate the differential impacts for distinct quantiles of the grade distribution.

Based upon conceptual frameworks outlined in Todd and Wolpin (2003) and Meghir and Rivkin (2011), the first estimation considers the academic achievement as the outcome of knowledge acquisition in a cumulative process, which is dependent on the individual endowment and a history of school and family inputs into the educational production function. Here $A_{i,t}$ is the academic achievement for a student i at time t and class j (constant group of students with whom the student takes all courses throughout high school):

$$A_{ij,t} = A_t\Big[F_i(t), S_{ij}(t), \mu_{i0}, \epsilon_{ij,t}\Big]$$

The vector of family inputs F_i is a function of time, as the family may commit different investments depending on the age of the child. The school inputs S_{ij} vary by the class j and may also be implemented in a different way by student i is also a function of time. The school inputs include the teachers and curriculum for the profiles (English, German, literature, etc.) as well as course schedule (morning-afternoon), varying by the class j. Here μ_{i0} denotes the student's individual endowment or capacity for cognitive achievement, which is time-invariant. The term $\epsilon_{ij,t}$ captures measurement error in the test scores.

i. Value-added and within-student estimation

The baseline specification estimates the average term GPA or the term grade obtained in a particular course (maths, literature, history, foreign languages) as a function of school inputs and controls and a partial adjustment mechanism from the lagged dependent variable:

$$GPA^k_{i,t} \quad = \quad \alpha + \beta Afternoon_{i,t} + \gamma GPA^k_{i,t-1} + \mathbf{x}_{i,t}\delta + \mathbf{w}_{i,t}\theta + \mu_i + \epsilon_{i,t} \tag{1}$$

Where the dependant variable $GPA^k_{i,t}$ denotes the grade of student i in subject k during term t (each school year has two academic terms). The indicator $Afternoon_{i,t}$ determines whether the student studied in the morning or the afternoon and β is the coefficient of interest, comparing average student performance between morning and afternoon classes holding other educational inputs constant. This coefficient should have a causal interpretation as students were allocated randomly to morning and afternoon schedules, without the possibility for self-selection.

$GPA^k_{i,t-1}$ is the student's previous term GPA, a lagged control for the individual academic achievement, a latent variable capturing previous school and family inputs in the educational production function. The full specification includes both the average term GPA (the mean of academic performance of subjects studied during the terms) as well as an autoregressive term, which is the GPA in the same

[15] Afternoon classes are not related to the profile of the student or their grade (Table A3).

subject, which is the dependent variable. Pre-determined covariates are contained in $\mathbf{w}_{i,t}$, they are contemporaneously exogenous with $\mathbb{E}[w_{i,t}\epsilon_{i,s}] \neq 0$ for $s < t$ but $\mathbb{E}[w_{i,t}\epsilon_{i,s}] = 0$ for $s \geq t$. The lagged average term GPA should be treated as a pre-determined variable, where the current period's errors are uncorrelated with lags of the average term GPA, but may be correlated with future values of this variable. This is because disturbance in the maths scores today will affect the average GPA in future periods.

The covariates contained in $\mathbf{x}_{i,t}$ include individual-level observed controls: gender and which profile group the student is enrolled in, where four binary variables indicate a different profile for German, history, English and English + Greek relative to the baseline of literature as a main subject. It also contains observables which only vary by year or term and an indicator for the second term of the school year. The teacher fixed effects indicate the profile group teacher, who takes responsibility for the group of students as a mentor (not the teacher of the subject k). Absence is controlling for the number of classes the student missed during the term, including both the excused and the unexcused number of absences.

This baseline specification relaxes the assumption that only contemporaneous inputs determine current cognitive achievement, allowing for modelling learning as a cumulative process. The unobserved heterogeneity, specific to an individual student μ_i, which should not be related to any of the school inputs which are observed in order for this specification to identify the true parameter of interest using OLS, but this assumption is not fulfilled when including an autoregressive term, which is by construction correlated with the unobserved individual effects.

A second specification sets $\gamma = 1$ and then the model can be rewritten as:

$$\Delta GPA^k_{i,t} = \alpha + \beta Afternoon_{i,t} + \mathbf{x}_{i,t}\delta + \mathbf{w}_{i,t}\theta + \mu_i + \epsilon_{i,t} \tag{2}$$

The difference in the test scores $\Delta GPA^k_{i,t} = GPA^k_{i,t} - GPA^k_{i,t-1}$ represents the value added as a function only of contemporaneous inputs. This formulation may be still very restrictive as the production function should not be age-varying, where both the inputs and the ability endowments should have the same impact regardless of age. The unobserved heterogeneity μ_i may be still related to the other educational inputs and the coefficients.

The third specification adds first differencing of the other observed educational inputs:

$$\Delta GPA^k_{i,t} = \alpha + \beta \Delta Afternoon_{i,t} + \Delta\mathbf{x}_{i,t}\delta + \Delta\mathbf{w}_{i,t}\theta + \epsilon_{i,t} \tag{3}$$

This estimator should remove μ_i - permanent factors specific to pupils, which are unobservable - so if an individual studied five years (ten school terms) in high school, the fixed unobservable would be differenced out within a student over time. Using the within-estimation imposes no restrictions about the distribution of the unobserved factors conditional on the included covariates. The unobserved academic capacity may be correlated with the observed school inputs.

The results present an alternative estimation of Equation (3) using fixed effects rather than first-differencing. There is an argument that fixed effects might be biased towards zero if the afternoon effect persists over periods, yet the lagged term average GPA should account for any accumulated impact over previous periods. The results will present both the first-difference and the fixed effects estimation for comparison. The inclusion of a lagged average term GPA, however, can be associated with introducing a potential bias and should be treated as a pre-determined regressor.

The full dynamic panel data model accounts for the bias introduced by the inclusion of the autoregressive term. Nickell (1981) shows that for small fixed total time periods and very large number of individuals in a panel, the fixed effects estimator does not produce consistent estimates. This results from the within transformation, which subtracts the average value from the dependent variable and regressors, inducing a correlation between the regressor and the error term. The bias on the autoregressive term is negative when there is positive autocorrelation in the dependent variable. Adding more covariates as controls would not remove the bias and if these covariates are correlated with the lagged dependent variable their coefficients would be also biased. The general form for p lags of the dependent variable included as covariates is:

$$GPA^k_{i,t} = \sum_{p=1}^{P} \gamma_p GPA^k_{i,t-p} + \beta Afternoon_{i,t} + \mathbf{x}_{i,t}\delta + \mathbf{w}_{i,t}\theta + \mu_i + \epsilon_{i,t} \tag{4}$$

To deal with this bias Anderson and Hsiao (1981); Arellano and Bond (1991); Holtz-Eakin et al. (1988) construct instruments for the endogenous lagged dependent variable from the previous lags in difference or level form. These instruments are uncorrelated with the composite error process (given the errors are *i.i.d.*) but are highly correlated with the lagged dependent variable. In the main table, I estimate the dynamic panel data model with the difference GMM where additional moment conditions resulting from the first-differencing of strictly exogenous variables enable more efficient estimation. In the robustness checks, I re-estimate using the system GMM approach outlined in Arellano and Bover (1995) and Blundell and Bond (1998), add further lags after conducting the Arellano-Bond test for autocorrelation and relax the assumption of no serial correlation, allowing for an MA(1) process in the error terms.

ii. Difference-in-difference: pre-reform and afternoon classes

Using the value-added and individual fixed effects specification, I consider both the sample pre-2012 as well as the the complete sample over the nine year period and model regime change. Note that there is a difference between morning classes during the double-shift system and morning classes post-regime change, in terms of i) start times: classes from 2012 onwards started 25 minutes later (7:45am rather than 7:20am), ii) classes duration: 45 minutes rather than 40 minutes and iii) end times (through later start times and longer classes): the school day could last until 3:15pm from 2012 onwards, i.e. a much longer school day, which extends past lunch. Therefore, morning classes pre- and post-2012 could be different.

In the following, I define three separate categories: (i) Morning classes pre-2012, (ii) Afternoon classes pre-2012 and (iii) Morning classes post-2012. Morning classes pre-2012 are the omitted baseline category and two indicators are included for afternoon classes and classes post-2012 respectively:

$$GPA^k_{i,t} = \sum_{p=1}^{P} \gamma_p GPA^k_{i,t-p} + \beta Afternoon_{i,t} + \kappa MorningPost_{i,t} + \mathbf{x}_{i,t}\delta + \mathbf{w}_{i,t}\theta + \mu_i + \epsilon_{i,t} \tag{5}$$

Here κ identifies the difference between academic performance in morning classes pre- and post-

2012. I consequently estimate the value-added and fixed effects of model equations (1) - (4) with the addition of the new indicator and including the entire 2008-2016 dataset.

Without observing academic performance during afternoon classes post-reform as the school schedule had only morning start times, there is no actual category to compare using difference-in-difference, but I use this estimator in a falsification test. This is based on the idea that if there is a true difference between student performance in morning and afternoon school schedules, this should be a systematic difference, while if we were to randomly assign students to morning or afternoon classes in a placebo way (post-reform all student only go to morning classes), then any difference from this placebo assignment should be non-systematic.

Therefore the falsification test difference-in-difference model simulates a random difference post-reform and compares it with the actual difference in morning and afternoon academic performance pre-reform. This model serves to set up a counterfactual difference, which is by construction non-systematic between academic terms of the school year post-regime change.

In the difference-in-difference specification, I define two indicators where $PseudoAfternoon_s$ is 1 for $Afternoon$ and 0 for $Morning$ before 2012 and is a randomly generated dummy, taking values 1 or 0 post-reform. Similarly, the indicator $Before$ takes the value of 1 if the observed value is pre-2012 when the school run the double-shift schedule of classes and 0 for $After$ when all classes were in the morning. The regression difference-in-difference form is then:

$$GPA_{i,t}^{k} = \alpha + \beta_1 PseudoAfternoon_s + \beta_2 Before_t + \beta_3 (PseudoAfternoon_s * Before_t) + ... + \epsilon_{i,ts} \tag{6}$$

The indices are t for time and s for the type of shift (morning / afternoon). This model is also subject to the same specifications, including the lagged term GPA, controlling for past accumulated achievement and student fixed effects. As all observations post the regime change would have only a pseudo dummy of shift (morning-afternoon) allocated to them, $PseudoAfternoon_s$ should not be related to the subject term GPAs. The parameter of interest here is β_3, which is the difference-in-difference regression coefficient identifying the causal impact of the afternoon school schedule in the quasi-experimental setting.

$$\begin{aligned}\beta_3 &= \mathbf{E}[Y_{ist}|s = Afternoon, t = Before] - \mathbf{E}[Y_{ist}|s = Morning, t = Before] \\ &- \{\mathbf{E}[Y_{ist}|s = Afternoon, t = After] - \mathbf{E}[Y_{ist}|s = Morning, t = After]\}\end{aligned}$$

The first term in the difference stands for the difference between morning and afternoon classes before regime change, before 2012. The second term in the difference represents a non-systematic difference in test scores between school terms during the post-regime change period, where ($s = Afternoon, t = After$) is placebo treatment. The β gives the parameter of interest as the difference-in-difference regression estimator.

iii. Quantile regression: differential impacts in the grade distribution

The analysis so far considers the impact of the randomly assigned afternoon classes through the estimation of the conditional mean function. In the next, the aim is to uncover differences to the treatment of afternoon classes and responses to the key covariates across different segments - and at the extremes - of the grade distribution. Following Koenker (2005); Koenker and Hallock (2001), I define the quantile regression model for the binary treatment problem of afternoon classes as the τ^{th} quantile of the distribution of the GPA in subject k (maths, history):

$$Q_{GPA_i^k}(\tau|D_i) = \alpha(\tau) + \delta(\tau)D_i$$

where the treatment indicator D_i stands or afternoon classes or morning classes post-2012 and $\delta(\tau)$ is the quantile treatment effect. Assuming the effect is linear and putting together the indicator treatment variables and all other controls in one set of covariates $\mathbf{x_i}$ including the continuous variables, the following general form of the quantile function becomes:

$$Q_{GPA_i^k}(\tau|\mathbf{x}_i) = \alpha(\tau) + \mathbf{x}_i\beta(\tau) \tag{7}$$

where $\mathbf{x}_i = Afternoon, MorningPost, Secondterm, Female, TermGPA_{t-1}, Absence$. The simultaneous quantile estimation also includes individual indicators for profiles, grades and years as controls. The obtained estimator $\hat{\beta}(\tau)$ minimises the asymmetric absolute loss function conditional on the covariates in $\mathbf{x}_i$.

iv. Within year variation between terms

This robustness check investigates to what extend there is higher variation between term grades pre-2012 (a student would have alternated between morning and afternoon classes) compared to post-2012. If afternoon classes have an impact on test scores then they would induce higher variation in the term average GPA compared to morning-only start times. The average term GPA exhibits slightly higher variation before 2012 (see Figure 5), though this is only pronounced for the cross-section of 9th graders). While the first-differencing expresses the gain or loss in subject term GPA in consecutive periods, in order to explore the within year variation in term grades, I define a difference between yearly and term subject grade. This specification uses the difference between the term average GPA in a subject and the overall year GPA in the same subject, which is the average of the two term GPAs for a subject k:

$$GPAgap_{i,t}^k = GPA_{i,t}^k - GPAyear_{i,t}^k$$

If the gap is positive, the student is over-performing in that term relative to the year average, if it is negative, then the student is under-performing in the specific subject. For a large part of the students there is no difference between academic performance in the two school terms (Figure 7 and Figure 6 in the Appendix), so the analysis aims to uncover whether the gap is systematically related to the shift pattern of morning and afternoon classes. The within-year variation isolates a difference in grades for a subject taught by the same teacher and having a structured curriculum specific to the school year. This variation can be related to the conditions the student experienced during the respective term in which one is over- or under-performing. Certainly whether it is the first or second term of the school year and the course schedule can have an impact on the variation of the grades within a year, so these covariate are included as controls. This new definition also allows to compare the gap between terms pre- and post regime change. If the afternoon classes were associated with significant difference between subject term scores, transitioning to a the same start times may be related to a decrease in the subject score gap between the two terms within the same year.

V. Results

i. Main results

Value-added and within student estimation: restricted and full sample

Table 2 summarises the results from the model equations (1) - (4) for the subjects mathematics and history, where the afternoon effect retains statistical significance in all specifications and robustness checks. Columns (1) and (2) show the OLS specification and Column (3) uses the added-value model with contemporaneous inputs. Columns (4) and (5) estimate the the first-difference and the fixed effects together with a lagged term average GPA score. Columns (6) - (7) use the difference GMM estimation for linear dynamic panel data (Arellano and Bond, 1991) with addition of the lagged average term GPA as a pre-determined covariate in the last column.

The estimate of the afternoon impact on academic performance in mathematics is a reduction of 0.074 (0.019) standard deviations. Conversely, the impact on the history GPA is positive: afternoon classes are associated with history scores higher by 0.064 (0.030) standard deviations in the full GMM model (significant at 5 %). The fixed effects model accounts for a slightly lower reduction in math scores and lower increase in history scores compared to the first-difference model, which contain only the lagged term average as a value-added control. The negative coefficient of the lagged average term GPA in the first-difference columns speak for potential Nickell bias for the pre-determined covariate. In the GMM models for both maths and history, the previous term average GPA is positively significant. The autoregressive term is, however only significant in the history GPA regression, yet the magnitude is larger for the lagged term average GPA. Note that the use of a dynamic panel data model - through instrumenting the lagged dependent variable and the pre-determined covariate - significantly reduces the number of observations in the already unbalanced panel of history GPA. History is not studied in all years and therefore there are gaps in T.

These results are consistent with previous findings[16] although the full model here is GMM as opposed to first-difference, where first-difference presents a higher magnitude of the afternoon affect compared to the dynamic panel data model. The same specifications applied to other subjects including literature, foreign languages and sports do not retain statistical significance in the full dynamic model, therefore there is no evidence for an afternoon effect for the other subjects (see Tables A4 and A5 in the Appendix). The coefficient on afternoon classes on term average GPA is positive and significant in some model specifications including the dynamic panel data model, but once other lagged subject GPAs are included as pre-determined covariate, it is no longer significant.

In the full sample the afternoon classes' effects are very similar, slightly larger in magnitude, to the restricted sample before regime change (see Table 3). Afternoon classes are associated with a reduction of 0.082 (0.018) standard deviations in the math GPA and an increase of 0.069 (0.029) in the history GPA. The indicator $MorningPost_{i,t}$ is positive and significant for both maths and history GPAs in the fixed effects specification, with a 0.737 (0.042) and 1.534 (0.068) standard deviations increase associated with morning classes post-2012 relative to morning classes pre-2012, respectively. There are several differences between morning classes pre- and post-regime change, one of which is later start times.

[16] Pope (2016) finds 0.072 (0.006) decrease in maths GPA associated with an afternoon instead of morning class using a first-difference model with contemporaneous inputs

Table 2: *Pre-reform sample (2008-2011): Panel data value-added, impact on math and history GPA from afternoon classes. Columns (1)-(2) use OLS, where (2) adds controls and an autoregressive term. Columns (3)-(4) uses first-difference specification where (3) has only contemporaneous inputs (levels) and (4) adds differencing of time-varying inputs between t and $t-1$, note the reported coefficients in (4) are for the differenced covariates. Column (5) presents the results from fixed effects within estimation and Columns (6)-(7) estimate Arellano-Bond dynamic panel with difference GMM, consequently adding also the previous term average GPA as a pre-determined covariate. Number of students in (6)-(7) are 729 for math GPA and 516 for history GPA. Instruments for the differenced equation GMM-type: second lag of MathGPA or HistoryGPA, respectively and first lag of $TermGPA_{t-1}$. All grades and absences are standardised with mean zero and variance one. Fixed effects include separate time-varying indicators for grade and year and time-invariant indicators for profile and teacher. Robust standard errors in brackets, clustered by student for (1)-(5) and robust standard errors in (6)-(7). Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
Math GPA	mathGPA	mathGPA	ΔmathGPA	ΔmathGPA	mathGPA	mathGPA	mathGPA
Afternoon	-0.052***	-0.117***	-0.193***	-0.094***	-0.067***	-0.088***	-0.074***
	[0.013]	[0.022]	[0.027]	[0.018]	[0.020]	[0.019]	[0.019]
Second term	0.001	0.098***	0.102***	0.076***	0.088***	0.083***	0.088***
	[0.014]	[0.026]	[0.032]	[0.019]	[0.022]	[0.019]	[0.019]
TermGPA_{t-1}		0.240***	-0.161***	-0.240***	0.063*		0.210***
		[0.020]	[0.012]	[0.033]	[0.033]		[0.066]
MathGPA_{t-1}		0.486***				0.019	-0.037
		[0.021]				[0.036]	[0.034]
Female	0.406***	0.090***	0.081***				
	[0.067]	[0.023]	[0.017]				
Ex. absence		-0.060***	-0.062***	-0.058***	-0.077***	-0.064***	-0.072***
		[0.012]	[0.011]	[0.016]	[0.018]	[0.017]	[0.017]
Unex. absence		-0.138***	-0.115***	-0.090***	-0.129***	-0.101***	-0.098***
		[0.018]	[0.018]	[0.023]	[0.026]	[0.024]	[0.024]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	4,913	3,915	3,915	2,924	3,915	2,924	2,924
R-squared	0.037	0.641	0.099	0.086	0.801		

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
History GPA	histGPA	histGPA	ΔhistGPA	ΔhistGPA	histGPA	histGPA	histGPA
Afternoon	0.102***	0.098***	0.184***	0.124***	0.119***	0.060**	0.064**
	[0.022]	[0.032]	[0.043]	[0.025]	[0.029]	[0.030]	[0.030]
Second term	0.187***	0.312***	0.469***	0.316***	0.312***	0.302***	0.280***
	[0.021]	[0.036]	[0.050]	[0.026]	[0.032]	[0.034]	[0.034]
TermGPA_{t-1}		0.349***	-0.199***	-0.310***	0.046		0.254**
		[0.031]	[0.020]	[0.053]	[0.054]		[0.115]
HistoryGPA_{t-1}		0.270***				0.150***	0.095**
		[0.028]				[0.057]	[0.048]
Female	0.333***	0.012	0.063**				
	[0.075]	[0.036]	[0.027]				
Ex. absence		-0.059***	-0.067***	-0.026	-0.039	-0.040	-0.051*
		[0.017]	[0.017]	[0.024]	[0.024]	[0.027]	[0.027]
Unex. absence		-0.102***	-0.101***	-0.037	-0.083**	-0.072*	-0.072*
		[0.027]	[0.028]	[0.031]	[0.039]	[0.038]	[0.038]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	3,269	2,498	2,498	2,069	2,834	1,734	1,734
R-squared	0.035	0.519	0.182	0.142	0.669		

This is some evidence that later start times for morning-only schedules are associated with better performance. However, the indicator for morning classes is not significant in the full dynamic model with further lags and accounting for serial correlation[17].

Grades in both subjects are significantly higher during the second term of the school year in both the restricted pre-2012 sample as well as the full sample estimation over the nine years. The second term difference is smaller in magnitude for mathematics compared to history: 0.088 (0.019) vs. 0.280 (0.034), respectively. The second term indicator is an important control because curriculum, learning dynamics, testing and grading may be substantially different in the second part of the school year. Reportedly, teachers may strategically choose to give lower grades in the first term to incentive students to work harder in the second term or apply more generous grading at the end of the year, compared to the first term. The higher grades can be also a consequence of accumulated learning and better performance on the part of the students in the second term.

Noting some of the covariates, the excused absence rate is negatively significant in most model specifications with 0.068 (0.011) decrease for maths GPA and 0.069 (0.017) decrease in history GPA associated with a standard deviation increase in the excused absences in the full sample over the nine years. The unexcused absences are negatively significant in the full dynamic models for maths scores in the restricted sample with a coefficient of 0.098 (0.024) standard deviations decrease, but become only marginally significant in the full sample. The opposite is valid for the history scores' association with unexcused absence: 0.029 (0.013) decrease in the full sample (significant at 5%).

Unexcused absences accumulate through being late for class but also skipping classes for non-legitimate reasons and are proxy for student behaviour. While the absence from class in itself is probably not a direct cause for the variation in grades, higher unexcused absence rates are associated with student behaviour patterns, which reflect student types, and attitudes towards the learning process. Unexcused absences are significantly higher during afternoon classes terms (see Table A3), which may indicate that students are more likely to skip classes during the school terms with afternoon classes.

Quantile regression: differential impacts in the grade distribution

The variation in academic performance is larger in profile groups of the subjects literature and history and (see Figure 3 in the Appendix). This again reflects the preference for foreign language profiles and the better academic achievement in the profile groups of English, German and English + Greek. The bulk of the distribution of maths, history, term average and year averages GPAs for students from the foreign language profiles English and German is towards the top of the grade scale (between "Very good" 5 and "Excellent" 6). For the distinct profile groups, the average math and history GPA split by morning and afternoon classes pre-2012 (see Figures 8 and 8) shows different patterns, where only the literature profile always had their average math GPA lower in afternoon classes terms and average history GPA higher in afternoon terms, respectively.

Tables 4 an 5 present the quantile regression results of history and maths GPA on the set of covariates and Figure 4 graphs the coefficients at different quantiles of the distribution. The negative impact of afternoon classes on the maths scores is largest for the lowest quantile: 0.128 (0.045) standard

[17] $MorningPost_{i,t}$ is not a significant indicator in the system GMM models, instrumenting for the autoregressive bias and accounting for serial correlation in the error terms (in a model with year trend and specified as a nodifference instrument). The robustness models in Tables **??** and **??** use year and grade fixed effects instead and do not report $MorningPost_{i,t}$ - which can be collinear with the year fixed effects, which capture better the differences between respective years for the main variables of interest.

Table 3: *Full sample (2008-2016): Panel data value-added, impact on math and history GPA from afternoon classes. Columns (1)-(2) use OLS, where (2) adds controls and an autoregressive term. Columns (3)-(4) uses first-difference specification where (3) has only contemporaneous inputs (levels) and (4) adds differencing of time-varying inputs between t and $t-1$, note the reported coefficients in (4) are for the differenced covariates. Column (5) presents the results from fixed effects within estimation and Columns (6)-(7) estimate Arellano-Bond dynamic panel with difference GMM, consequently adding also the previous term average GPA as a pre-determined covariate. Number of students in (6)-(7) are 1,325 for math GPA and 1,088 for history GPA. Instruments for the differenced equation GMM-type: second lag of MathGPA or HistoryGPA, respectively and first lag of $TermGPA_{t-1}$. All grades and absences are standardised with mean zero and variance one. Fixed effects include separate time-varying indicators for grade and year and time-invariant indicators for profile and teacher. Robust standard errors in brackets, clustered by student. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
Math GPA	mathGPA	mathGPA	ΔmathGPA	ΔmathGPA	mathGPA	mathGPA	mathGPA
Afternoon	-0.052***	-0.129***	-0.213***	-0.096***	-0.085***	-0.094***	-0.082***
	[0.013]	[0.021]	[0.025]	[0.018]	[0.019]	[0.018]	[0.018]
MorningPost	-0.083**	0.070**	0.094***		0.737***	0.000	0.000
	[0.038]	[0.035]	[0.035]		[0.042]	[0.000]	[0.000]
Second term	0.011	0.092***	0.102***	0.086***	0.087***	0.086***	0.085***
	[0.009]	[0.016]	[0.020]	[0.011]	[0.013]	[0.012]	[0.012]
$TermGPA_{t-1}$		0.275***	-0.122***	-0.226***	0.147***		0.157***
		[0.014]	[0.008]	[0.021]	[0.019]		[0.040]
$MathGPA_{t-1}$		0.496***				0.018	-0.026
		[0.014]				[0.023]	[0.022]
Female	0.317***	0.064***	0.042***				
	[0.053]	[0.016]	[0.009]				
Ex. absence		-0.069***	-0.071***	-0.062***	-0.077***	-0.066***	-0.068***
		[0.008]	[0.007]	[0.010]	[0.010]	[0.011]	[0.011]
Unex. absence		-0.038*	-0.041*		-0.032	-0.031*	-0.030*
		[0.022]	[0.022]	[0.015]	[0.022]	[0.017]	[0.016]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	10,861	9,214	9,214	7,601	9,215	7,600	7,600
R-squared	0.022	0.669	0.066	0.070	0.800		

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
History GPA	histGPA	histGPA	ΔhistGPA	ΔhistGPA	histGPA	histGPA	histGPA
Afternoon	0.103***	0.123***	0.207***	0.128***	0.110***	0.072**	0.069**
	[0.022]	[0.031]	[0.040]	[0.026]	[0.029]	[0.030]	[0.029]
MorningPost	-0.059	0.331***	0.403***		1.534***	0.000	0.000
	[0.044]	[0.060]	[0.064]		[0.068]	[0.000]	[0.000]
Second term	0.152***	0.247***	0.371***	0.209***	0.215***	0.222***	0.208***
	[0.014]	[0.022]	[0.031]	[0.015]	[0.018]	[0.020]	[0.020]
$TermGPA_{t-1}$		0.381***	-0.155***	-0.261***	0.167***		0.195***
		[0.019]	[0.012]	[0.032]	[0.033]		[0.069]
$HistoryGPA_{t-1}$		0.289***				0.104***	0.058**
		[0.019]				[0.031]	[0.027]
Female	0.253***	-0.023	0.063***				
	[0.055]	[0.025]	[0.016]				
Ex. absence		-0.091***	-0.097***	-0.046***	-0.082***	-0.063***	-0.069***
		[0.011]	[0.011]	0.015]	[0.015]	[0.017]	[0.017]
Unex. absence		-0.034**	-0.035**	-0.024**	-0.028**	-0.029**	-0.029**
		[0.013]	[0.017]	[0.010]	[0.011]	[0.014]	[0.013]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	6,963	5,595	5,595	5,153	6,499	4,252	4,252
R-squared	0.023	0.516	0.113	0.081	0.663		

Table 4: *Quantile regression estimates for maths and history GPAs, full sample 2008-2016. All columns contain fixed effects for grade, profile and year. Simultaneous-quantile regression results in all columns. Bootstrap standard errors in brackets. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Maths GPA	q10	q20	q30	q40	q50	q60	q70	q80	q90
Afternoon	-0.128***	-0.079**	-0.068**	-0.072***	-0.057***	-0.057***	-0.063***	-0.059**	-0.033
	[0.045]	[0.034]	[0.033]	[0.018]	[0.014]	[0.018]	[0.018]	[0.028]	[0.023]
MorningPost	-0.148**	-0.138***	-0.104**	-0.045	-0.049**	-0.031	-0.040	-0.048	-0.007
	[0.075]	[0.043]	[0.051]	[0.037]	[0.024]	[0.033]	[0.035]	[0.055]	[0.035]
Second term	0.081***	0.077***	0.066***	0.068***	0.075***	0.072***	0.069***	0.050***	0.047***
	[0.028]	[0.024]	[0.017]	[0.011]	[0.011]	[0.011]	[0.010]	[0.016]	[0.018]
Female	0.090***	0.113***	0.119***	0.045**	0.045***	0.047***	0.028	0.025	0.041*
	[0.026]	[0.016]	[0.024]	[0.019]	[0.016]	[0.015]	[0.021]	[0.023]	[0.021]
TermGPA$_{t-1}$	0.649***	0.686***	0.741***	0.759***	0.735***	0.706***	0.661***	0.588***	0.436***
	[0.020]	[0.012]	[0.012]	[0.015]	[0.014]	[0.012]	[0.013]	[0.012]	[0.022]
Ex. absence	-0.078***	-0.080***	-0.076***	-0.053***	-0.055***	-0.052***	-0.046***	-0.054***	-0.052***
	[0.014]	[0.014]	[0.013]	[0.009]	[0.009]	[0.009]	[0.008]	[0.007]	[0.008]
Unex. absence	-0.089***	-0.097***	-0.066**	-0.050	-0.039	-0.031	-0.024	-0.030	-0.011
	[0.025]	[0.025]	[0.031]	[0.031]	[0.030]	[0.030]	[0.030]	[0.027]	[0.022]
Observations	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215	9,215
Pseudo R^2	0.2911	0.3728	0.3699	0.4114	0.3711	0.3657	0.3377	0.2130	0.0454

Table 5: *Quantile regression estimates for maths and history GPAs, full sample 2008-2016. All columns contain fixed effects for grade, profile and year. Simultaneous-quantile regression results in all columns. Bootstrap standard errors in brackets. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
History GPA	q10	q20	q30	q40	q50	q60	q70
Afternoon	0.073	0.084	0.065*	0.073***	0.072***	0.069***	0.066***
	[0.052]	[0.054]	[0.036]	[0.025]	[0.023]	[0.019]	[0.022]
MorningPost	0.203**	0.124	0.028	0.056	0.049	0.081***	0.039
	[0.086]	[0.100]	[0.080]	[0.050]	[0.042]	[0.026]	[0.029]
Second term	0.251***	0.250***	0.201***	0.148***	0.130***	0.123***	0.107***
	[0.053]	[0.036]	[0.027]	[0.018]	[0.016]	[0.014]	[0.016]
Female	-0.005	-0.042	-0.034	-0.047**	-0.044**	-0.036**	-0.029
	[0.046]	[0.039]	[0.025]	[0.020]	[0.019]	[0.014]	[0.018]
TermGPA_{t-1}	0.753***	0.796***	0.767***	0.720***	0.630***	0.523***	0.423***
	[0.023]	[0.021]	[0.017]	[0.013]	[0.014]	[0.010]	[0.016]
Excused absence	-0.131***	-0.114***	-0.090***	-0.059***	-0.042***	-0.044***	-0.040***
	[0.021]	[0.019]	[0.013]	[0.013]	[0.012]	[0.009]	[0.009]
Unexcused absence	-0.101**	-0.075***	-0.048**	-0.041***	-0.044***	-0.034***	-0.031**
	[0.041]	[0.025]	[0.021]	[0.015]	[0.016]	[0.013]	[0.012]
Observations	6,499	6,499	6,499	6,499	6,499	6,499	6,499
Pseudo R^2	0.3105	0.3504	0.2858	0.2917	0.3020	0.1819	0.0532

deviations decrease related to afternoon classes and it gradually decreases for higher quantiles of the distribution. The magnitude for the median of the maths grade distribution is a decrease of 0.057 (0.014) standard deviations for afternoon classes. In contrast the OLS estimation with full controls showed a 0.075 (0.018) standard deviations decrease for the mean of the distribution. This suggests that students with lower grades already stand to lose more by having maths classes scheduled in the afternoon.

The distribution of maths GPA is more spread out than the history GPA. More than half of all history term grades in the full sample are the maximum achievable score of "Excellent" 6 (see Figure 2), which complicates estimations of marginal quantile effects at higher quantiles of the history grade distribution. The bottom of Table **??** presents the results of the quantile regression for history GPA up to the 70th quantile of the distribution. Statistically significant afternoon effect is only observed between the 40th and the 70th quantile of the distribution with declining magnitude for higher quantiles. The median afternoon effect is an increase of 0.073 (0.025) standard deviations.

There is statistically significant negative association between the excused absences and the maths as well as the history grades, where the magnitude of this relationship decreases for higher quantiles of the grade distribution. The same is also observed for the relationship between unexcused absences and history grades. However, unexcused absences are only significantly related to math grades in the lower quantiles of the distribution. Arguably, this is also evidence that accumulating unexcused absences through skipping class is a proxy for behaviours of students with the lowest academic achievement.

ii. Robustness and falsification tests

Tables 6 and 7 present the results of further dynamic panel data analysis including specifications outlined in Arellano and Bond (1991); Arellano and Bover (1995) and Blundell and Bond (1998). For math GPA, the coefficient on afternoon classes is almost the same in the Arellano-Bover/Blundell-Bond

(1995, 1998) model which uses lagged levels as well as lagged differences as instruments: a reduction of 0.081 (0.020) standard deviations. The difference in the afternoon effect on history GPA is larger between the difference and the system GMM, where the AB/BB model estimates an increase of 0.091 (0.032) standard deviations.

The Arellano-Bond test for zero autocorrelation in the first-differenced errors shows that the null hypothesis of zero autocorrelation can be rejected in up to the second order at 1% and at the third order at 10%. Further specifications in Table 6 include lags of the autoregressive math GPA and the pre-determined covariate term average GPA from the previous three periods. For history, only the first order autoregression test is significant at 1% and the third order is significant at 10 %. Given the unbalanced history scores panel and the result from the post-estimation test, I retain the first order autoregressive term in later specifications for history GPA.

Adding the further lags slightly increases the magnitude of the afternoon effect to a reduction of 0.090 (0.029) standard deviations. In later specifications, the term average GPA retains positive significance, even where the autoregressive term is no longer significant. This speaks for the use of the average term GPA as a pre-determined covariate, which captures better the total value added.

Columns (4)-(7) introduce more flexible syntax allowing for serial correlation in the disturbances, using higher order lags for the differenced composite errors. If the disturbances follow an MA(1) process then the composite errors would contain a lagged element with $\mu_i + \lambda\epsilon_{i,t-1} + \epsilon_{i,t}$, then second or higher lags of the math scores will be valid instruments for the level equation. For the differenced composite errors $\Delta\lambda\epsilon_{i,t-1} + \Delta\epsilon_{i,t}$ only third and further lags would be appropriate instruments. Starting with the third lag as an instrument for the difference equation and the second lag for the level equation in Column (4), I increase the lags in subsequent columns and consider the Sargan test of overidentifying restrictions with the null hypothesis that the instruments are valid[18]. More distant level lags actually increase the Sargan test statistic, therefore Column (7) uses the eighth difference lag and the third level lag as GMM instruments.

Accounting for possible serial correlation in the error terms and the clustering of the errors by student results in the dynamic panel data setting increases the magnitude of the afternoon effect. The estimate of the afternoon effect on math GPA is a reduction of 0.139 (0.042) standard deviations and for history GPA, afternoon classes are associated with an increase of 0.083 (0.038) standard deviations (significant at 5%). There is no significant difference between morning class before and after the regime change. Notably, other covariates also remain significant in the full dynamic models. The second term is associated with 0.084 (0.022) lower math scores and 0.187 (0.029) higher history scores. While unexcused absence is only marginally associated with lower history scores, excused absence is negatively significant for both subjects in all dynamic specifications. This indicates that being present in the classroom may be important, yet the relationship is not identified, as excused absence accumulates through sickness, which may be adversely affecting academic outcomes.

Table 8 presents the results from the difference-in-difference model with the full sample, which serves as a falsification test. The $PseudoAfternoon_s$ is not significant as expected, given that all values from 2012 onwards are randomly generated. The difference-in-difference regression coefficient $PseudoAfternoon_s * Before_t$, which identifies the effect of afternoon classes is significant for both maths and history test scores in all OLS, FD, FE and GMM models which also account for serial correlation.

[18] Yet, as Roodman (2009) points out, the Sargan test is not very powerful when there are many instruments.

The magnitude of the difference-in-difference estimator is a reduction in math GPA of 0.150 (0.051) and increase in history GPA of 0.144 (0.059) standard deviations, respectively. This magnitude increases with using further lags. These results identify the morning-afternoon difference as a systematic difference in test scores.

Figure 6 shows that for the majority of students there is no gap in academic performance between the two academic terms within a year with morning and afternoon classes, respectively (most of observations at zero). Morning classes are associated with slightly larger positive gap for mathematics and slightly larger negative gap for history and the reverse is true for afternoon classes with respect to both subjects. Figure 7 also demonstrates that post-2012 the gap in performance between term and year average is more frequently zero than before 2012, which indicates that there was a larger variation in test scores before the transition to morning-only start times.

The regression results in Tables A6 and A7 demonstrate the impact of afternoon classes on the math and history GPA gap between term and year for the full sample. In the simple OLS model with controls, the afternoon classes are associated with 0.079 (0.015) standard deviations negative gap in the math GPA and 0.119 (0.023) standard deviations positive gap in the history GPA. These magnitudes are slightly lower than those identified in Tables 6 and 7.

VI. Interpretation and Discussion

The exogenous variation resulting from random assignment to morning and afternoon class times creates an opportunity to investigate the differential time-of-day effects on student performance in a range of subjects. The magnitudes of the afternoon effect are not small and constitute nearly half of the smaller class size effect of about 0.2 standard deviations for allocating students to smaller classes (Krueger et al., 1999). Changes to the school schedule, however, do not require significant investment of additional resources: altering school programmes of scheduling mathematics courses earlier in the day and other courses later in the day may require only some school administration and adaptation costs, with significant benefits as a result.

The aim of this analysis was to uncover the impact of afternoon classes and whether the afternoon effect varies in different subjects. There is a limitation to the external validity of these findings, as the context is a high school, which specialises in humanities and foreign languages. Given that afternoon classes negatively impact math test scores, these impacts may be even more significant for schools, which are more reliant on math and science subjects. However, the magnitude of the uncovered effects from this dataset is similar to what other recent studies report. Substantial differences in curriculum, cognitive tasks and learning dynamics exist between high school subjects where mathematics is the most analytical extreme, based on problem-solving, which can be repetitive and automatised, requiring a higher degree of concentration and mental focus. Literature and languages are subjects in the middle of the range, where studying vocabulary and the structure of the language may require analytical thinking, yet employ some degree of narrative. On the other extreme, history is a subject, which employs a high degree of narrative. The findings indicate that significant results exist only for mathematics and history: subjects on the opposite ends of different mental task requirements.

The panel data structure allows for eliminating the individual specific time-invariant heterogeneities and as students do not select into either morning or afternoon classes, the uncovered effects are average for morning and afternoon chronotypes. Even if some students would be disproportionately affected by

Table 6: *Robustness check for the impact of afternoon classes on math GPA in the full sample (2008-2016). Column (1) coincides with the baseline results in Column (7) of Table 3 in the main results using a difference GMM as proposed by Arellano-Bond (1991). Columns (2)-(3) use the Arellano-Bover/Blundell-Bond (1995, 1998) estimator instead. After the postestimation tests Column (3) adds additional lags. Columns (4)-(7) use the more flexible syntax of the linear dynamic panel data estimation (LDPD) to relax the assumption of no serial correlation and use higher lags for the differenced composite errors. The rows "dgmmiv" and "lgmmiv" give the number of lags used for instrumenting the differenced and the level equation, respectively. The table also contains the results from the postestimation Sargan test of overidentifying restrictions testing for the validity of the instruments. All columns use robust standard errors, reported in brackets. Columns (4)-(7) use standard errors adjusted for clustering on the student level using the two-step estimator as described in Windmeijer (2005). Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	AB	AB/BB	AB/BB	LDPD	LDPD	LDPD	LDPD
MathGPA	MathGPA	MathGPA	MathGPA	MathGPA	MathGPA	MathGPA	MathGPA
Afternoon	-0.082***	-0.081***	-0.090***	-0.074**	-0.093***	-0.142***	-0.139***
	[0.018]	[0.020]	[0.029]	[0.034]	[0.033]	[0.037]	[0.042]
Second term	0.085***	0.087***	0.090***	0.089***	0.094***	0.091***	0.084***
	[0.012]	[0.013]	[0.018]	[0.016]	[0.017]	[0.020]	[0.022]
Ex. absence	-0.068***	-0.078***	-0.088***	-0.092***	-0.085***	-0.087***	-0.088***
	[0.011]	[0.012]	[0.014]	[0.014]	[0.014]	[0.015]	[0.013]
Unex. absence	-0.030*	-0.031*	-0.025	-0.026	-0.028	-0.025	-0.023
	[0.016]	[0.017]	[0.016]	[0.029]	[0.031]	[0.024]	[0.018]
MathGPA_{t-1}	-0.026	0.040*	-0.097**	-0.181**	-0.203**	0.004	-0.101
	[0.022]	[0.021]	[0.047]	[0.081]	[0.096]	[0.112]	[0.108]
MathGPA_{t-2}			0.048	0.107***	0.114	0.067	0.116
			[0.043]	[0.036]	[0.078]	[0.106]	[0.113]
MathGPA_{t-3}			-0.000	0.032	-0.003	-0.034	0.056*
			[0.025]	[0.029]	[0.086]	[0.111]	[0.034]
TermGPA_{t-1}	0.157***	0.213***	0.699***	0.379***	0.425***	0.362**	0.360**
	[0.040]	[0.032]	[0.079]	[0.084]	[0.108]	[0.149]	[0.162]
TermGPA_{t-2}			0.011	0.177***	0.159*	0.134	0.311
			[0.088]	[0.061]	[0.093]	[0.134]	[0.194]
TermGPA_{t-3}			0.014	0.030	0.036	0.069	-0.069
			[0.050]	[0.033]	[0.078]	[0.102]	[0.095]
dgmmiv				L(3/.)	L(5/.)	L(8/.)	L(8/.)
lgmmiv				L(2/.)	L(2/.)	L(2/.)	L(3/.)
Sargan test				chi2(186)	chi2(130)	chi2(58)	chi2(56)
				297.658	218.153	102.324	75.657
				0.0000	0.0000	0.0003	0.0412
Observations	7,600	9,214	6,286	6,286	6,286	6,286	6,286
# students	1,325	1,600	1,325	1,325	1,325	1,325	1,325

Arellano-Bond test for zero autocorrelation in first-differenced errors	Order	z	Prob > z
implemented after Column (2).	1	-24.623	0.0000
	2	4.4495	0.0000
	3	-1.8223	0.0684
	4	-.13782	0.8904
	5	.59715	0.5504

Table 7: *Robustness check for the impact of afternoon classes on history GPA in the full sample (2008-2016). Column (1) coincides with the baseline results in Column (7) of Table 3 in the main results using a difference GMM as proposed by Arellano-Bond (1991). Column (2) uses the Arellano-Bover/Blundell-Bond (1995, 1998) estimator instead. Columns (3)-(6) use the more flexible syntax of the linear dynamic panel data estimation (LDPD) to relax the assumption of no serial correlation and use higher lags for the differenced composite errors.The rows "dgmmiv" and "lgmmiv" give the number of lags used for instrumenting the differenced and the level equation, respectively. The table also contains the results from the postestimation Sargan test of overidentifying restrictions testing for the validity of the instruments. All columns use robust standard errors, reported in brackets. Columns (3)-(6) use standard errors adjusted for clustering on the student level using the two-step estimator as described in Windmeijer (2005). Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variable:	AB	AB/BB	LDPD	LDPD	LDPD	LDPD
History GPA	HistGPA	HistGPA	HistGPA	HistGPA	HistGPA	HistGPA
Afternoon	0.069**	0.091***	0.065**	0.075**	0.092**	0.083**
	[0.029]	[0.032]	[0.032]	[0.034]	[0.040]	[0.038]
Second term	0.208***	0.214***	0.158***	0.169***	0.193***	0.187***
	[0.020]	[0.022]	[0.025]	[0.025]	[0.028]	[0.029]
Ex. absence	-0.069***	-0.084***	-0.074***	-0.074***	-0.071***	-0.075***
	[0.017]	[0.018]	[0.021]	[0.020]	[0.021]	[0.022]
Unex. absence	-0.029**	-0.025**	-0.031	-0.032	-0.018*	-0.021
	[0.013]	[0.012]	[0.021]	[0.022]	[0.009]	[0.013]
HistoryGPA$_{t-1}$	0.058**	0.104***	-0.036	-0.036	0.136	0.208*
	[0.027]	[0.027]	[0.088]	[0.094]	[0.118]	[0.119]
TermGPA$_{t-1}$	0.195***	0.392***	0.559***	0.555***	0.396***	0.536***
	[0.069]	[0.045]	[0.089]	[0.101]	[0.111]	[0.101]
dgmmiv			L(3/.)	L(5/.)	L(8/.)	L(8/.)
lgmmiv			L(2/.)	L(2/.)	L(2/.)	L(3/.)
Sargan test			chi2(175)	chi2(117)	chi2(49)	chi2(47)
			296.997	205.857	90.706	86.260
			0.0000	0.0000	0.0003	0.0004
Observations	4,252	5,595	5,595	5,595	5,595	5,595
# students	1,088	1,336	1,336	1,336	1,336	1,336

Arellano-Bond test for zero autocorrelation in first-differenced errors	Order	z	Prob > z
implemented after Column (2).	1	-18.976	0.0000
	2	.91319	0.3611
	3	1.7416	0.0816
	4	-.61007	0.5418
	5	-.47152	0.6373

Table 8: *Difference-in-difference falsification test for maths and history GPA as set out in equation 6 using the full sample (2008-2016). Column (1) uses OLS, Columns (2) and (3) first difference and fixed effects, respectively. Columns (4)-(6) add autoregressive component with dynamic panel data using the estimator of Arellano-Bover/Blundell-Bond (1995, 1998), specifying three lags for math GPA and one lag for history GPA as in Tables 6 and 7 (only the first-order lag reported in the table). Columns (4)-(6) use the more flexible syntax of the linear dynamic panel data estimation (LDPD) to relax the assumption of no serial correlation and use higher lags for the differenced composite errors. The rows "dgmmiv" and "lgmmiv" give the number of lags used for instrumenting the differenced and the level equation, respectively. A year trend and grade fixed effects used all columns. Number of student_id for GMM models in math sub-table: 1,325 and in history sub-table 1,336. Columns (4)-(6) use standard errors adjusted for clustering on the student level using the two-step estimator as described in Windmeijer (2005). Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variable:	OLS	FD	FE	GMM	GMM	GMM
Math GPA	mathGPA	ΔmathGPA	mathGPA	mathGPA	mathGPA	mathGPA
PseudoAfternoon	-0.011	-0.021	-0.027*	0.011	0.015	0.014
	[0.017]	[0.015]	[0.015]	[0.024]	[0.025]	[0.026]
Before	-0.040	0.009	0.056**	0.020	0.054	0.190***
	[0.030]	[0.036]	[0.028]	[0.061]	[0.063]	[0.069]
PseudoAfternoon * Before	-0.064**	-0.077***	-0.052**	-0.150***	-0.174***	-0.236***
	[0.025]	[0.023]	[0.024]	[0.051]	[0.052]	[0.057]
Secondterm	0.081***	0.047***	0.088***	0.070***	0.080***	0.063***
	[0.014]	[0.010]	[0.013]	[0.015]	[0.016]	[0.022]
Excused absence	-0.068***	-0.063***	-0.077***	-0.092***	-0.087***	-0.083***
	[0.011]	[0.010]	[0.010]	[0.015]	[0.014]	[0.015]
Unexcused absence	-0.029	-0.028*	-0.032	-0.026	-0.027	-0.024
	[0.019]	[0.015]	[0.023]	[0.029]	[0.032]	[0.020]
MathGPA_{t-1}				-0.209**	-0.243**	-0.063
				[0.089]	[0.100]	[0.126]
TermGPA_{t-1}	0.665***	-0.221***	0.149***	0.426***	0.505***	0.422***
	[0.014]	[0.021]	[0.019]	[0.088]	[0.100]	[0.145]
dgmmiv				L(3/.)	L(5/.)	L(8/.)
lgmmiv				L(2/.)	L(2/.)	L(3/.)
Observations	9,215	7,601	9,215	6,286	6,286	6,286
R-squared	0.565	0.063	0.798			

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variable:	OLS	FD	FE	GMM	GMM	GMM
History GPA	histGPA	ΔhistGPA	histGPA	histGPA	histGPA	histGPA
PseudoAfternoon	-0.000	0.004	0.010	-0.043	-0.037	-0.076*
	[0.024]	[0.024]	[0.026]	[0.038]	[0.035]	[0.043]
Before	-0.028	-0.354***	-0.058	-0.249***	-0.348***	-0.100
	[0.047]	[0.071]	[0.058]	[0.082]	[0.087]	[0.142]
PseudoAfternoon * Before	0.122***	0.132***	0.081**	0.144**	0.162***	0.221***
	[0.037]	[0.036]	[0.040]	[0.059]	[0.055]	[0.065]
Second term	0.174***	0.054***	0.188***	0.158***	0.147***	0.168***
	[0.017]	[0.015]	[0.018]	[0.025]	[0.025]	[0.035]
Excused absence	-0.092***	-0.047***	-0.086***	-0.064***	-0.075***	-0.074***
	[0.013]	[0.015]	[0.016]	[0.020]	[0.020]	[0.025]
Unexcused absence	-0.039***	-0.024**	-0.031**	-0.022	-0.024	-0.020
	[0.014]	[0.010]	[0.013]	[0.017]	[0.017]	[0.014]
TermGPA_{t-1}	0.590***	-0.271***	0.163***	0.568***	0.598***	0.605***
	[0.015]	[0.032]	[0.033]	[0.088]	[0.101]	[0.160]
HistoryGPA_{t-1}				-0.019	-0.079	0.112
				[0.088]	[0.097]	[0.208]
dgmmiv				L(3/.)	L(5/.)	L(8/.)
lgmmiv				L(2/.)	L(2/.)	L(3/.)
Observations	6,499	5,153	6,499	5,595	5,595	5,595
R-squared	0.431	0.076	0.650			

having the class scheduled earlier or later during the day, the differencing over time should eliminate individual differences for a person overtime. However, there are differences in the quantile estimates: students from the lower grade distribution stand to lose more by having maths classes scheduled in the afternoon, while the same students may not benefit by having history classes in the afternoon. The winners and losers of having optimal or counterproductive scheduling of classes, respectively, are not evenly spread in the grade distribution.

Alternative confounding factors may relate not simply to the performance and mental capacity differences in student between morning and afternoon classes, but also to the way teachers organise their day and the constraints a double-shift schedule puts on them. Having morning-only classes enables teachers to have their lectures in one block with fewer interruptions than a double-shift schedule where they would spend longer working days in the school with larger gaps in their teaching. While students with afternoon classes would arrive directly from home for their classes, teachers may have already spent the whole morning in the school. Longer working days - even with the same number of lectures - may induce additional fatigue for teachers in the afternoons. If such an effect exists, however, it should work in opposite ways for maths and history teachers, as history teachers should become systematically better in the afternoons than in the mornings (or there is a self-selection of teachers with evening chronotypes in the history profession). For teachers themselves, the time-of-day effects may work in similar ways and would have an indirect impact on students' test scores.

VII. Conclusion

Using a decade long dataset from a high school, which instituted a double-shift schedule with alternation between morning and afternoon classes, I investigate the impact of afternoon classes on student performance in a range of subjects. The variation in scheduling was exogenously imposed with no possibility of self-selection into a certain school start time. The panel data with long time dimension allow for the estimation of a value-added dynamic panel data, differencing out individual specific time-invariant heterogeneities over time.

The findings indicate that afternoon classes lowered math test scores and increased history test scores, which relate to psychology and neuroscience research about optimal functioning in different times-of-the-day. These results have implications for optimal course scheduling and present a low-cost intervention, which can lead to increase in student performance without substantial investments in school inputs.

References

Anderson, T. W. and C. Hsiao (1981): "Estimation of dynamic models with error components," *Journal of the American statistical Association*, 76, 598–606.

Angrist, J. D. and V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *The Quarterly Journal of Economics*, 114, 533–575.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2013): "Explaining charter school effectiveness," *American Economic Journal: Applied Economics*, 5, 1–27.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2016): "Teacher Quality and Learning Outcomes in Kindergarten," *The Quarterly Journal of Economics*, 175–214.

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, 58, 277–297.

Arellano, M. and O. Bover (1995): "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68, 29–51.

Barnes, C. M. and D. T. Wagner (2009): "Changing to daylight saving time cuts into sleep and increases workplace injuries." *Journal of Applied Psychology*, 94, 1305.

Blatchford, R. (2015): "Schools must consider double shifting to accommodate growing pupil numbers," *The Guardian: Teacher Network, School leadership and management*.

Blundell, R. and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115–143.

Bray, M. (2008): *Double-shift schooling: Design and operation for cost-effectiveness*, vol. 90, Commonwealth Secretariat.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor supply of New York City cabdrivers: One day at a time," *The Quarterly Journal of Economics*, 407–441.

Card, D. and A. B. Krueger (1992): "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *The Journal of Political Economy*, 100, 1–40.

Cardinali, D. P. (2008): "Chronoeducation: How the biological clock influences the learning process," *The educated brain: Essays in neuroeducation*, 110–26.

Carrell, S. E., T. Maghakian, and J. E. West (2011): "A's from Zzzz's? The causal effect of school start time on the academic achievement of adolescents," *American Economic Journal: Economic Policy*, 3, 62–81.

Carskadon, M. A., C. Vieira, and C. Acebo (1993): "Association between puberty and delayed phase preference," *Sleep*, 16, 258–258.

Carskadon, M. A., A. R. Wolfson, C. Acebo, O. Tzischinsky, R. Seifer, et al. (1998): "Adolescent sleep patterns, circadian timing, and sleepiness at a transition to early school days," *Sleep*, 21, 871–881.

CHEN, D. L., T. J. MOSKOWITZ, AND K. SHUE (2016): "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires," *The Quarterly Journal of Economics*, 131, 1181–1242.

CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014): "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *The American Economic Review*, 104, 2633–2679.

CHOI, E. J., H. R. MOON, AND G. RIDDER (2014): "Estimation of an education production function under random assignment with selection," *The American Economic Review*, 104, 206–211.

CORTES, K. E., J. BRICKER, C. ROHLFS, ET AL. (2012): "The Role of Specific Subjects in Education Production Functions: Evidence from Morning Classes in Chicago Public High Schools," *The BE Journal of Economic Analysis & Policy*, 12, 1–36.

CRAIG, A. AND R. CONDON (1985): "Speed-accuracy trade-off and time of day," *Acta psychologica*, 58, 115–122.

CUNHA, F. AND J. HECKMAN (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78, 883–931.

DANZIGER, S., J. LEVAV, AND L. AVNAIM-PESSO (2011): "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences*, 108, 6889–6892.

DAS, J., S. DERCON, J. HABYARIMANA, P. KRISHNAN, K. MURALIDHARAN, AND V. SUNDARARAMAN (2013): "School inputs, household substitution, and test scores," *American Economic Journal: Applied Economics*, 5, 29–57.

DEMING, D. J., J. S. HASTINGS, T. J. KANE, AND D. O. STAIGER (2014): "School choice, school quality, and postsecondary attainment," *The American economic review*, 104, 991–1013.

DILLS, A. K. AND R. HERNANDEZ-JULIAN (2008): "Course scheduling and academic performance," *Economics of Education Review*, 27, 646–654.

DING, W. AND S. F. LEHRER (2014): "Understanding the role of time-varying unobserved ability heterogeneity in education production," *Economics of Education Review*, 40, 55–75.

DOBBIE, W. AND R. G. FRYER (2013): "Getting beneath the veil of effective schools: Evidence from New York City," *American Economic Journal: Applied Economics*, 5, 28–60.

DOLEAC, J. L. AND N. J. SANDERS (2015): "Under the cover of darkness: How ambient light influences criminal activity," *Review of Economics and Statistics*, 97, 1093–1103.

EDWARDS, F. (2012): "Early to rise? The effect of daily start times on academic performance," *Economics of Education Review*, 31, 970–983.

FARBER, H. S. (2015): "Why you Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers," *The Quarterly Journal of Economics*, 130, 1975–2026.

FOLKARD, S. AND P. TUCKER (2003): "Shift work, safety and productivity," *Occupational Medicine*, 53, 95–101.

GENTZKOW, M. AND J. M. SHAPIRO (2008): "Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study," *The Quarterly Journal of Economics*, 279–323.

GIBSON, M. AND J. SHRADER (2015): "Time Use and Productivity: The Wage Returns to Sleep," Tech. rep., Department of Economics, Williams College.

GOLDSTEIN, D., C. S. HAHN, L. HASHER, U. J. WIPRZYCKA, AND P. D. ZELAZO (2007): "Time of day, intellectual performance, and behavioral problems in morning versus evening type adolescents: Is there a synchrony effect?" *Personality and Individual Differences*, 42, 431–440.

HAMERMESH, D. S., C. K. MYERS, AND M. POCOCK (2008): "Cues for Timing and Coordination: Latitude, Letterman, and Longitude," *Journal of Labor Economics*, 26, 223–246.

HINRICHS, P. (2011): "When the bell tolls: The effects of school starting times on academic achievement," *Education Finance and Policy*, 6, 486–507.

HOLTZ-EAKIN, D., W. NEWEY, AND H. S. ROSEN (1988): "Estimating vector autoregressions with panel data," *Econometrica: Journal of the Econometric Society*, 1371–1395.

HWANG, H.-S., W. R. REED, AND C. HUBBARD (1992): "Compensating wage differentials and unobserved productivity," *Journal of Political Economy*, 835–858.

KAMSTRA, M. J., L. A. KRAMER, AND M. D. LEVI (2000): "Losing sleep at the market: The daylight saving anomaly," *The American Economic Review*, 90, 1005–1011.

KOENKER, R. (2005): *Quantile regression*, 38, Cambridge University Press.

KOENKER, R. AND K. HALLOCK (2001): "Quantile regression: An introduction," *Journal of Economic Perspectives*, 15, 43–56.

KOSTIUK, P. F. (1990): "Compensating differentials for shift work," *Journal of Political Economy*, 1054–1075.

KRUEGER, A. B. ET AL. (1999): "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 114, 497–532.

LANFRANCHI, J., H. OHLSSON, AND A. SKALLI (2002): "Compensating wage differentials and shift work preferences," *Economics Letters*, 74, 393–398.

LAVY, V. (2009): "Performance pay and teachers' effort, productivity, and grading ethics," *The American Economic Review*, 99, 1979–2021.

LAZEAR, E. P. (2001): "Educational Production," *The Quarterly Journal of Economics*, 116, 777–803.

LUSHER, L. AND V. YASENOV (2016): "Double-shift schooling and student success: Quasi-experimental evidence from Europe," *Economics Letters*, 139, 36–39.

MACKENBERG, E. J., D. M. BROVERMAN, W. VOGEL, AND E. L. KLAIBER (1974): "Morning-to-afternoon changes in cognitive performances and in the electroencephalogram." *Journal of Educational Psychology*, 66, 238.

MEGHIR, C. AND S. RIVKIN (2011): "Econometric methods for research in education," *Handbook of the Economics of Education*, 3, 1–87.

Miller, L. S., T. Lombardo, and S. Fowler (1992): "Time of day effects on a human force discrimination task," *Physiology & Behavior*, 52, 839–841.

Monk, T. H. and V. C. Leng (1982): "Time of day effects in simple repetitive tasks: some possible mechanisms," *Acta Psychologica*, 51, 207–221.

Nickell, S. (1981): "Biases in dynamic models with fixed effects," *Econometrica*, 1417–1426.

Pagel, J. F., N. Forister, and C. Kwiatkowki (2007): "Adolescent sleep disturbance and school performance: the confounding variable of socioeconomics." *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 3, 19–23.

Pope, N. G. (2016): "How the Time of Day Affects Productivity: Evidence from School Schedules," *Review of Economics and Statistics*, 98, 1–11.

Rice, J. K., R. G. Croninger, and C. F. Roellke (2002): "The effect of block scheduling high school mathematics courses on student achievement and teachers' use of time: implications for educational productivity," *Economics of Education Review*, 21, 599–607.

Roodman, D. (2009): "A note on the theme of too many instruments," *Oxford Bulletin of Economics and statistics*, 71, 135–158.

Rothstein, J. (2010): "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *The Quarterly Journal of Economics*, 125, 175–214.

Schilbach, F., H. Schofield, and S. Mullainathan (2016): "The Psychological Lives of the Poor," *The American Economic Review*, 106, 435–440.

Song, J. and C. Stough (2000): "The relationship between morningness–eveningness, time-of-day, speed of information processing, and intelligence," *Personality and Individual Differences*, 29, 1179–1190.

Todd, P. E. and K. I. Wolpin (2003): "On the specification and estimation of the production function for cognitive achievement," *The Economic Journal*, 113, F3–F33.

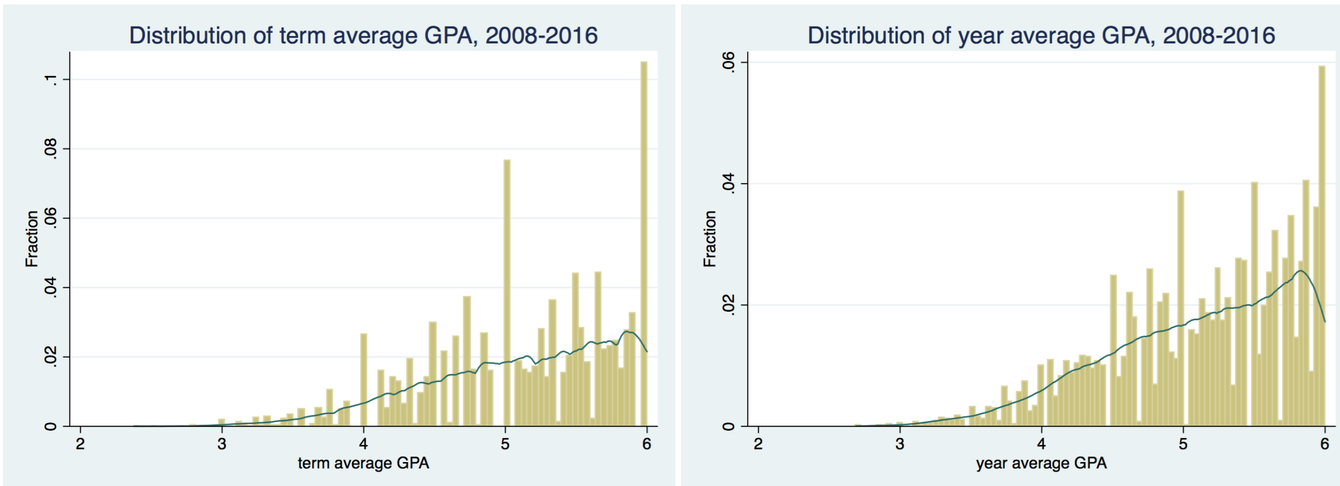
Ulmer, C., D. M. Wolman, M. M. Johns, et al. (2009): *Resident duty hours: enhancing sleep, supervision, and safety*, National Academies Press.

Windmeijer, F. (2005): "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of Econometrics*, 126, 25–51.

Wolfson, A. R. and M. A. Carskadon (1998): "Sleep schedules and daytime functioning in adolescents," *Child Development*, 69, 875–887.

Appendix

Figure 1: *Distribution of term average and year average GPA*

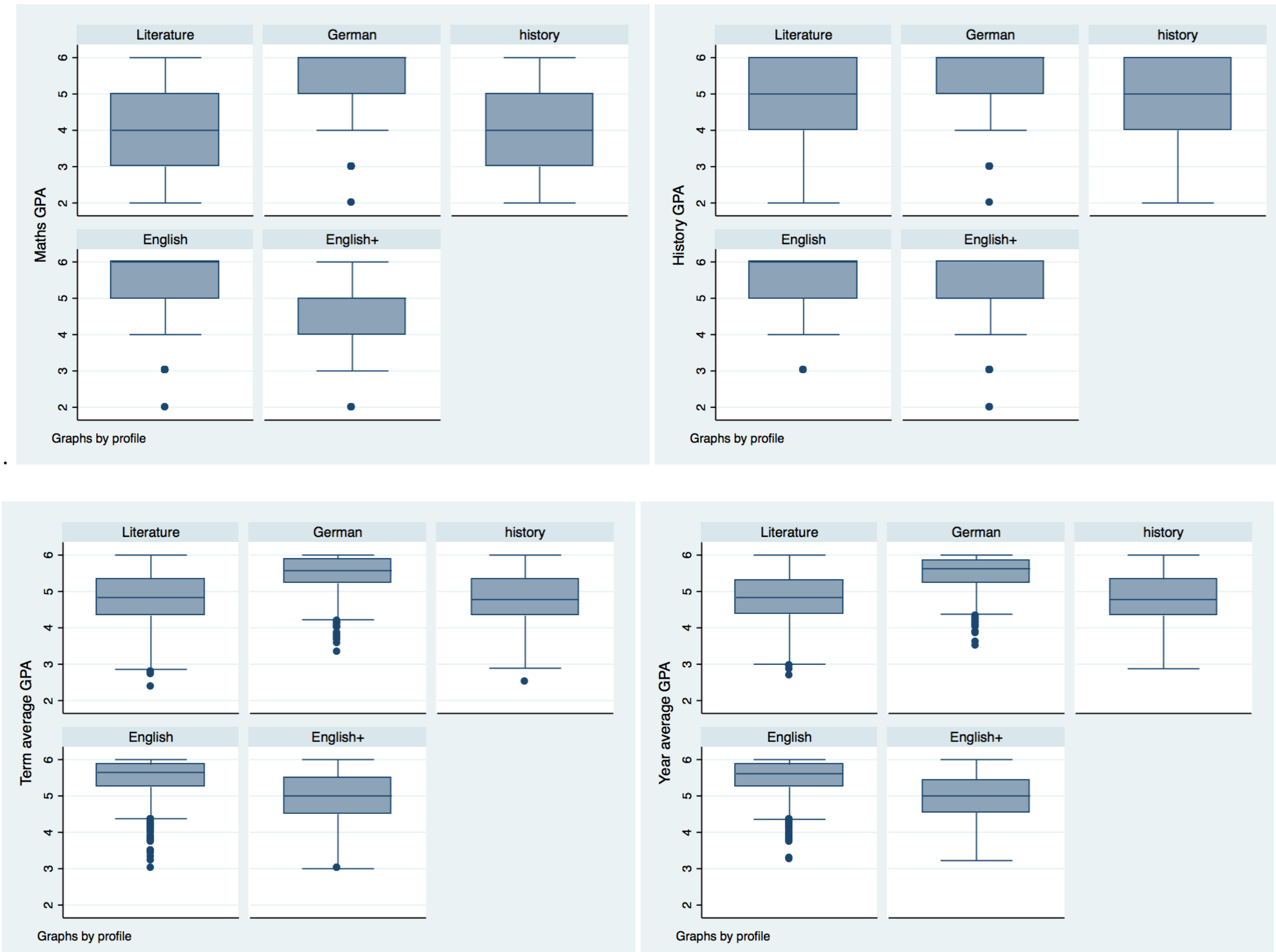


.

Figure 2: *Maths and history GPA distribution. Absolute values: 2 (poor) to 6 (excellent)*

.

Figure 3: *Box graph distribution of maths, history, term and year average GPA by profile group*



Horizontal bars in the middle of each box represent the median. The boxes' upper and lower limits are the first and the third quartile of the grade distribution. The bounds of the box whiskers are: the upper adjacent value is $x_{[75]} + \frac{3}{2}(x_{[75]} - x_{[25]})$ and the lower adjacent value is $x_{[25]} - \frac{3}{2}(x_{[75]} - x_{[25]})$, respectively.

Figure 4: *The coefficients of the quantile regressions of maths and history GPAs*

.

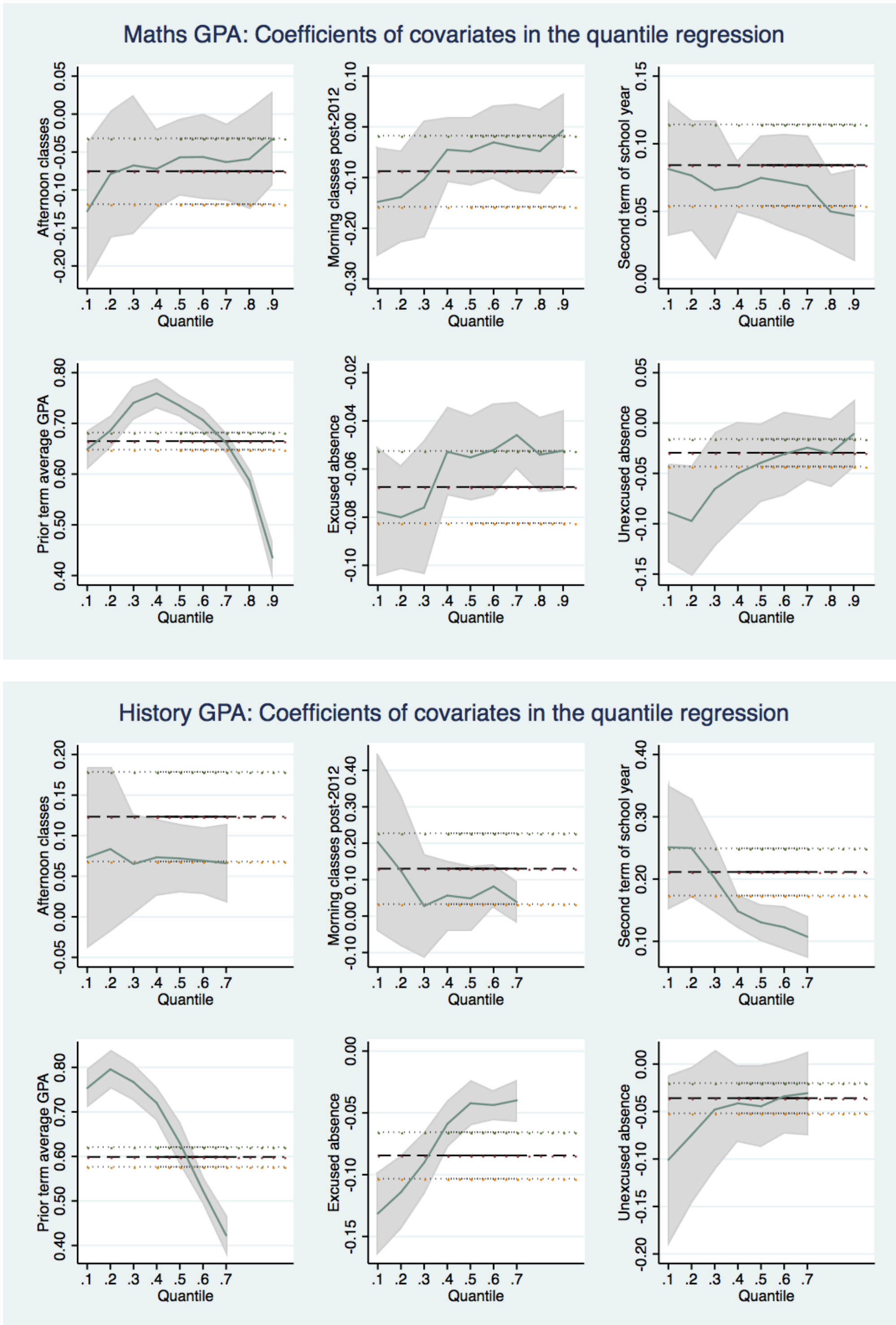


Figure 5: *Average term GPA by term and grade*

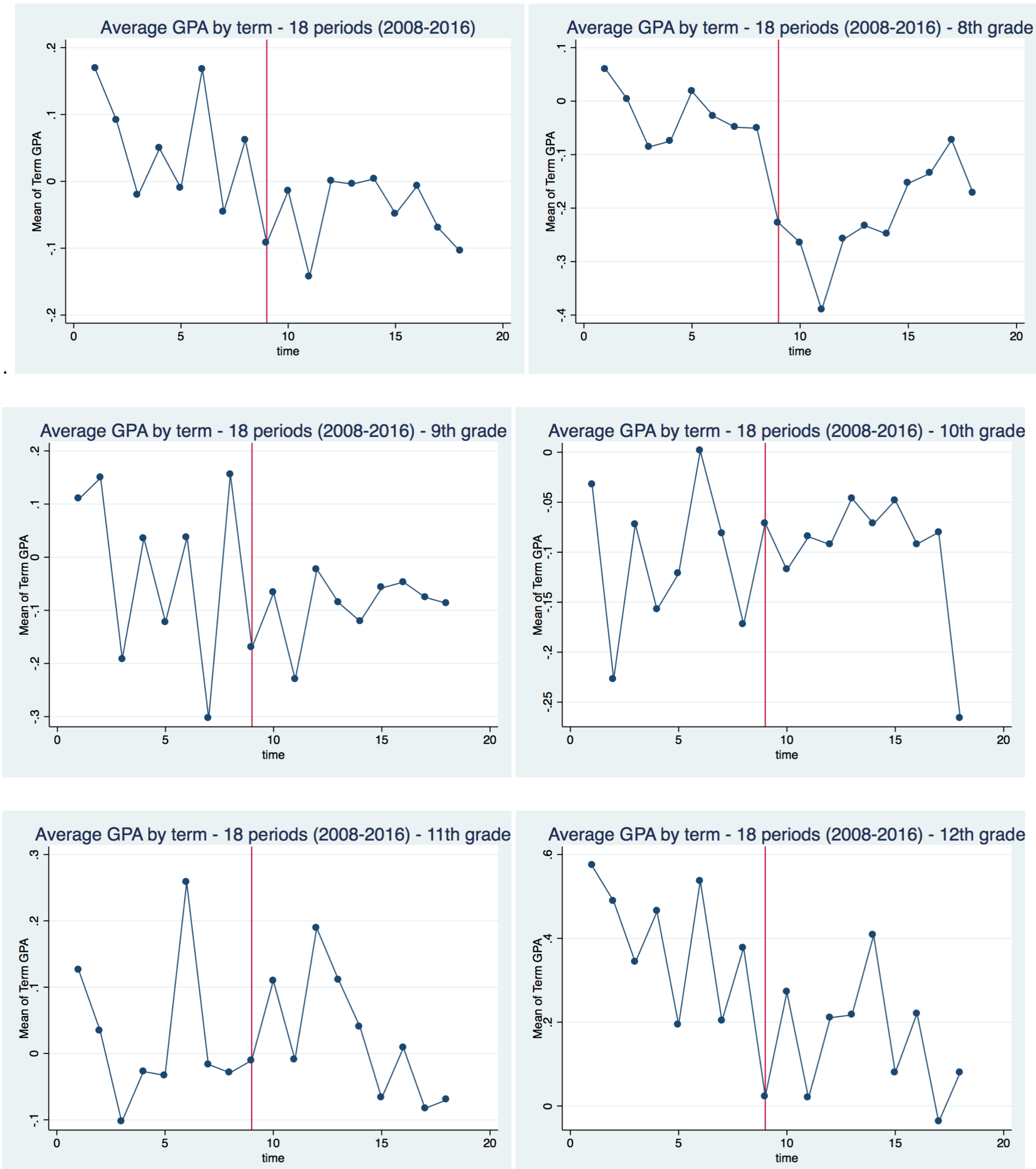
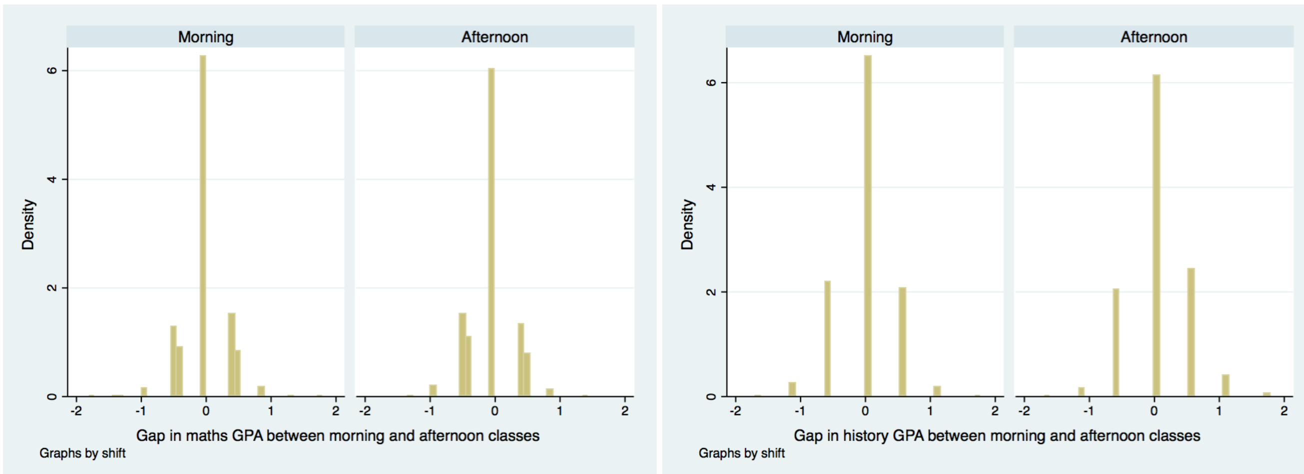
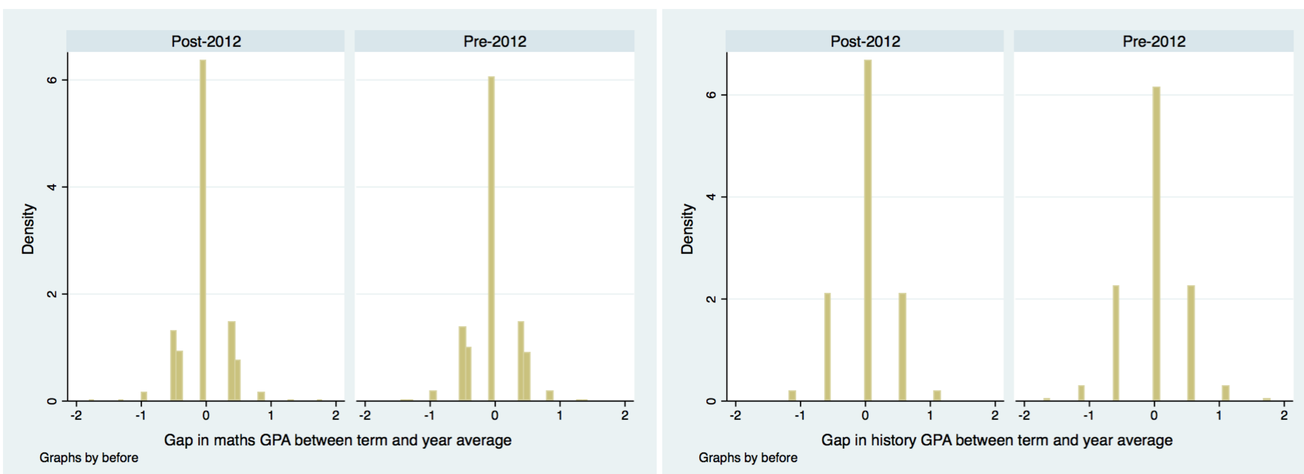


Figure 6: *Term-year gap in maths and history GPAs by morning and afternoon classes*



.

Figure 7: *Term-year gap in maths and history GPAs by period: post- and pre-2012*



.

Figure 8: *Average maths and history GPA. Pre-2012 average is split into morning and afternoon classes, by all profiles, literature and German profile*

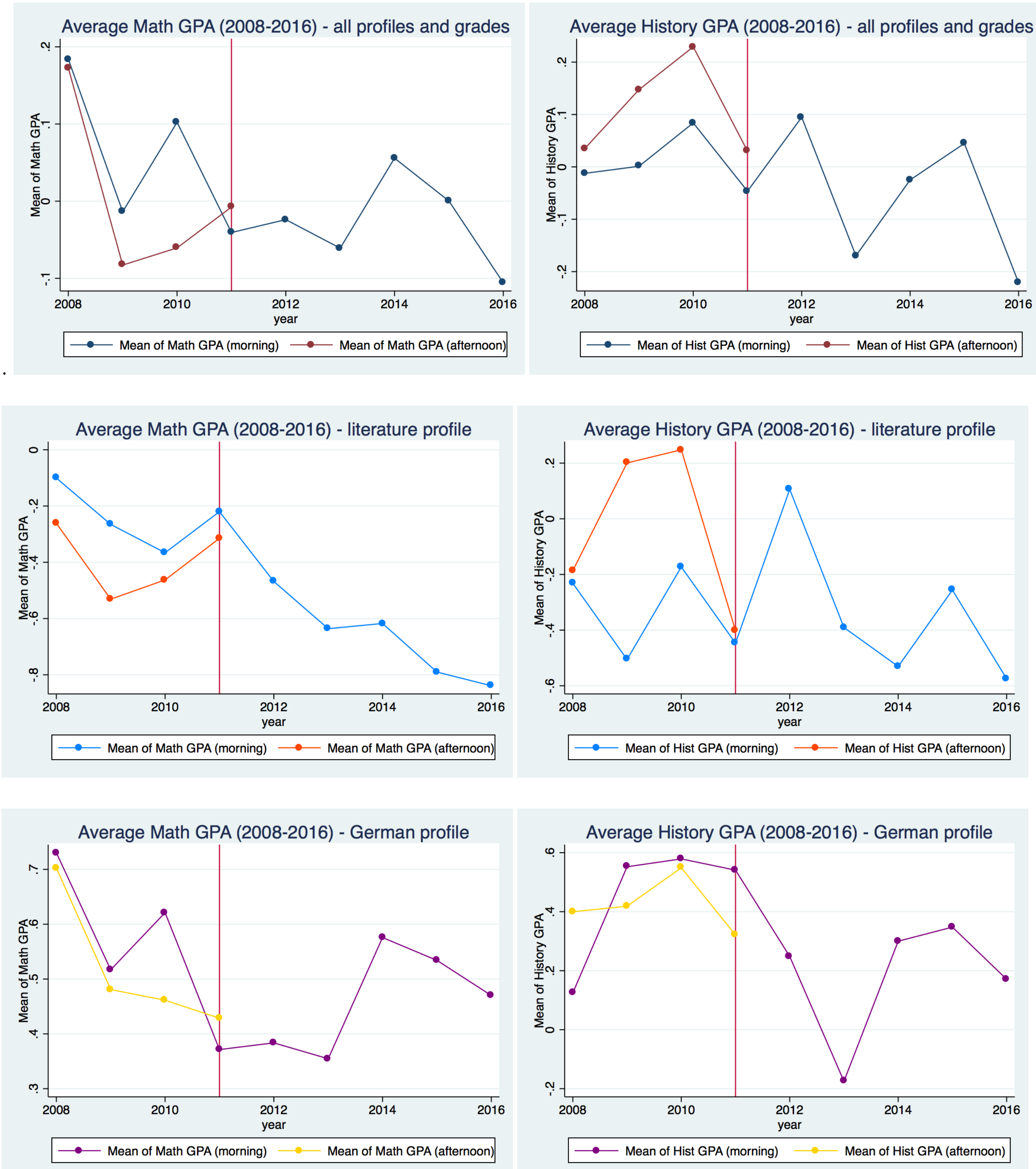


Figure 9: *Average maths and history GPA. Pre-2012 average is split into morning and afternoon classes, by history, English and English+ profile*

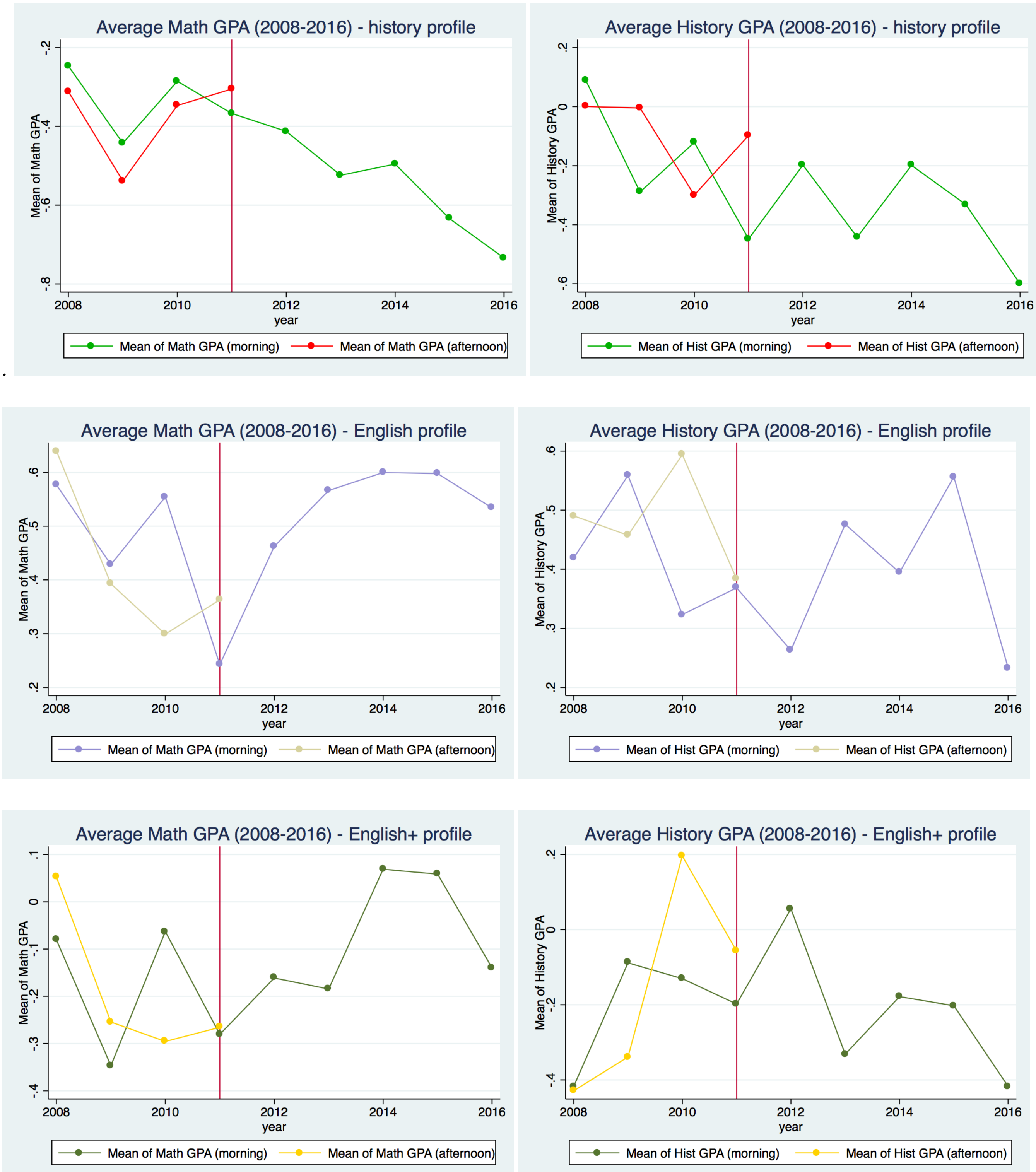


Table A1: *Summary statistics by sample: of overall term and subjects GPAs, final exam results, entry grades and absences. All grades GPAs and the final exams are measured on a scale from 2 to 6 (poor to excellent). The entry GPA is formed of the nationally held examination results in literature for the literature and history profiles and literature and maths examinations for the language profiles English, German and English + Greek as well as the 7th grade graduation diploma from middle school.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full sample	**Pre-reform**	**Post-reform**	**Morning**	**Afternoon**	**Pre-reform**	**Full sample**
	2008-2016	**2008-2011**	**2012-2012**	**2008-2011**	**2008-2011**	**2008-2011**	**2008-2016**
	N=10866	**N=4918**	**N=5948**	**N=2459**	**N=2459**	**N=5948**	**N=10866**
	Mean	**Mean**	**Mean**	**Mean**	**Mean**	**Difference**	**Difference**
						Morning-Afternoon	**After-Before**
overall term GPA	5.13	5.17	5.10	5.16	5.18	-0.014	-0.071
	(0.67)	(0.65)	(0.68)	(0.67)	(0.64)	(0.019)	(0.013)
literature GPA	5.29	5.37	5.22	5.36	5.37	-0.003	-0.147
	(0.75)	(0.72)	(0.77)	(0.72)	(0.71)	(0.020)	(0.014)
maths GPA	4.78	4.82	4.75	4.84	4.79	0.058	-0.066
	(1.11)	(1.07)	(1.14)	(1.07)	(1.07)	(0.031)	(0.021)
foreign language GPA	4.94	4.98	4.91	4.97	4.99	-0.012	-0.073
	(0.97)	(0.98)	(0.97)	(0.98)	(0.97)	(0.029)	(0.020)
history GPA	5.29	5.34	5.24	5.29	5.38	-0.092	-0.097
	(0.89)	(0.87)	(0.9)	(0.91)	(0.83)	(0.030)	(0.021)
sports GPA	5.73	5.82	5.67	5.81	5.82	-0.012	-0.150
	(0.54)	(0.44)	(0.6)	(0.45)	(0.43)	(0.013)	(0.011)
final exam literature	5.38	5.45	5.29	5.44	5.44	-	-0.155
	(0.51)	(0.48)	(0.53)	(0.48)	(0.48)	-	(0.011)
final exam option	5.20	5.17	5.23	5.17	5.17	-	0.060
	(0.76)	(0.76)	(0.77)	(0.76)	(0.76)	-	0.017
entry GPA	29.51	29.14	29.71	29.14	29.14	-	0.563
	(2.58)	(2.48)	(2.6)	(2.48)	(2.48)	-	(0.057)
excused absences	48.18	45.17	50.68	46.09	44.25	1.842	5.513
	(46.79)	(46.16)	(47.17)	(47.14)	(45.15)	(1.317)	(0.902)
non-excused absences	1.81	1.90	1.74	1.82	1.98	-0.158	-0.158
	(3.42)	(2.40)	(4.07)	(2.54)	(2.25)	(0.069)	(0.066)
female	0.71	0.72	0.71	0.72	0.72	-	-0.008
	(0.45)	(0.45)	(0.45)	(0.45)	(0.45)	-	(0.009)

Notes: Mean values of variables in full samples and subsamples in Columns 1-5. Standard deviations in paretheses. Columns 6 and 7 give the difference between morning and afternoon in the pre-reform subsample and the difference after-before in for the full sample with standard errors of the difference in parentheses.

Table A2: *Summary statistics by profile in the full sample: overall term and subjects GPAs, final exam results, entry grades and absences. All grades GPAs and the final exams are measured on a scale from 2 to 6 (poor to excellent). The entry GPA is formed of the nationally held examination results in literature for the literature and history profiles and literature and maths examinations for the language profiles English, German and English + Greek as well as the 7th grade graduation diploma from middle school.*

Profile:	**(a) Literature N=2092**	**(b) German N=2274**	**(c) History N=2074**	**(d) English N=2294**	**(e) English+ N=2132**
overall term GPA	4.80	5.50	4.81	5.50	4.98
	(0.68)	(0.47)	(0.68)	(0.47)	(0.62)
literature GPA	5.15	5.49	5.09	5.54	5.11
	(0.79)	(0.62)	(0.83)	(0.59)	(0.78)
maths GPA	4.22	5.33	4.26	5.33	4.64
	(1.13)	(0.83)	(1.13)	(0.85)	(1.02)
foreign language GPA	4.51	5.44	4.55	5.25	4.7
	(1.01)	(0.71)	(1.05)	(0.79)	(0.94)
history GPA	5.06	5.55	5.07	5.65	5.10
	(0.99)	(0.72)	(0.94)	(0.61)	(0.92)
sports GPA	5.61	5.84	5.63	5.82	5.74
	(0.64)	(0.43)	(0.61)	(0.43)	(0.54)
final exam literature	5.24	5.51	5.22	5.62	5.29
	(0.56)	(0.43)	(0.55)	(0.33)	(0.51)
final exam option	4.89	5.38	4.99	5.61	5.05
	(0.76)	(0.77)	(0.78)	(0.45)	(0.77)
entry GPA	27.83	31.16	27.86	31.65	28.69
	(2.16)	(2.02)	(2.10)	(1.51)	(2.06)
excused absences	53.19	45.53	47.87	46.74	47.96
	(48.37)	(45.1)	(44.4)	(46.03)	(49.67)
non-excused absences	2.18	1.27	2.25	1.63	1.81
	(2.85)	(1.85)	(3.99)	(5.05)	(2.08)
female	0.77	0.71	0.67	0.70	0.72
	(0.42)	(0.45)	(0.47)	(0.46)	(0.45)

Notes: Mean values by profile for the full sample 2007-2016 in Columns a-e. Standard deviations in parentheses. Sample size for an individual subject might be smaller if e.g. history is not studied in all years.

Table A3: *Descriptive statistics: profiles and grades coefficients on the likelihood of afternoon classes, female as well as OLS of average term GPA and subjects GPAs in maths and history on profiles and grades. Absences rates regressions on covarates.*

	(1) Probit Afternoon	(2) Probit Female	(3) OLS term GPA	(4) OLS math GPA	(5) OLS hist GPA
German	-0.034	-0.182	1.040***	1.004***	0.550***
	[0.053]	[0.117]	[0.067]	[0.065]	[0.066]
History	-0.001	-0.281**	0.003	0.040	0.016
	[0.054]	[0.119]	[0.078]	[0.076]	[0.078]
English	-0.032	-0.206*	1.048***	1.008***	0.669***
	[0.053]	[0.117]	[0.067]	[0.066]	[0.063]
English + Greek	-0.021	-0.157	0.269***	0.382***	0.049
	[0.054]	[0.120]	[0.075]	[0.071]	[0.076]
Grade 9	0.005	-0.022	0.064***	0.043**	0.822
	[0.022]	[0.020]	[0.019]	[0.020]	[0.672]
Grade 10	0.012	0.000	0.025	0.049**	0.797
	[0.030]	[0.028]	[0.025]	[0.024]	[0.674]
Grade 11	0.026	-0.015	0.154***	0.117***	0.933
	[0.036]	[0.034]	[0.027]	[0.026]	[0.674]
Grade 12	0.028	-0.010	0.392***	0.136***	1.007
	[0.041]	[0.038]	[0.027]	[0.028]	[0.673]
Year trend		-0.004	-0.024***	-0.019***	-0.032***
		[0.011]	[0.007]	[0.006]	[0.008]
Observations	10,880	10,880	10,866	10,861	6,963
Log likelihood	-5817.1246	-6496.1971	-	-	-
Pseudo R^2 / R-squared	0.0002	0.0039	0.253	0.203	0.099

Robust standard errors in brackets, clustered by student.
Statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Absence rate:	(1) Excused 2008-2011	(2) Excused 2008-2016	(3) Unexcused 2008-2011	(4) Unexcused 2008-2016
Afternoon	-0.015	-0.016	0.058***	0.055***
	[0.022]	[0.021]	[0.018]	[0.018]
Before		1.425***		0.189***
		[0.041]		[0.024]
Term GPA	-0.327***	-0.365***	-0.265***	-0.242***
	[0.041]	[0.025]	[0.029]	[0.018]
Secondterm	0.338***	0.315***	0.124***	0.068***
	[0.023]	[0.014]	[0.018]	[0.016]
Year FE	x	x	x	x
Grade FE	x	x	x	x
Teacher FE	x	x	x	x
Observations	4,912	10,828	4,912	10,834
R-squared	0.616	0.606	0.560	0.310

Robust standard errors in brackets, clustered by student.
Statistical significance: *** p<0.01, ** p<0.05, * p<0.1

Table A4: *Pre-reform sample (2008-2011): Panel data value-added, impact on term average GPA and literature GPA from afternoon classes. Columns (1)-(2) use OLS, where (2) adds controls and an autoregressive term. Columns (3)-(4) uses first-difference specification where (3) has only contemporaneous inputs (levels) and (4) adds differencing of time-varying inputs between t and $t-1$, note the reported coefficients in (4) are for the differenced covariates. Column (5) presents the results from fixed effects within estimation and Columns (6)-(7) estimate Arellano-Bond dynamic panel with difference GMM, consequently adding also the previous term average GPA as a pre-determined covariate in the literature table and previous maths, literature and foreign language GPA in the average term GPA table, respectively. Number of students in (6)-(7) are 730-681 for term GPA and 730 for literature GPA. Instruments for the differenced equation GMM-type: second lag of TermGPA or LiteratureGPA, respectively and first lag of the pre-determined covariates. All grades and absences are standardised with mean zero and variance one. Fixed effects include separate time-varying indicators for grade and year and time-invariant indicators for profile and teacher. Robust standard errors in brackets, clustered by student for (1)-(5) and robust standard errors in (6)-(7). Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
Term GPA	termGPA	termGPA	ΔtermGPA	ΔtermGPA	termGPA	termGPA	termGPA
Afternoon	0.025***	0.022	0.046***	0.008	-0.006	0.036**	0.019
	[0.009]	[0.016]	[0.018]	[0.012]	[0.014]	[0.015]	[0.014]
Second term	0.069***	0.156***	0.149***	0.123***	0.139***	0.169***	0.137***
	[0.009]	[0.018]	[0.019]	[0.010]	[0.014]	[0.015]	[0.014]
TermGPA$_{t-1}$		0.702***				0.452***	0.374***
		[0.023]				[0.048]	[0.049]
LiteratureGPA$_{t-1}$				-0.090***	-0.013		-0.067
				[0.012]	[0.016]		[0.021]
MathGPA$_{t-1}$				-0.085***	0.038**		-0.045
				[0.014]	[0.017]		[0.026]
ForeignGPA$_{t-1}$				-0.039***	0.050***		0.068
				[0.013]	[0.017]		[0.023]
Female	0.498***	0.112***	0.107***				
	[0.073]	[0.016]	[0.014]				
Ex.absence		-0.076***	-0.075***	-0.059***	-0.076***	-0.082***	-0.077***
		[0.008]	[0.008]	[0.011]	[0.013]	[0.015]	[0.015]
Unex. absence		-0.120***	-0.093***	-0.063***	-0.127***	-0.093***	-0.088***
		[0.014]	[0.013]	[0.015]	[0.020]	[0.019]	[0.018]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	4,918	3,574	3,574	2,644	3,574	2,928	2,644
R-squared	0.054	0.836	0.214	0.186	0.910		

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
Literature GPA	litGPA	litGPA	ΔlitGPA	ΔlitGPA	litGPA	litGPA	litGPA
Afternoon	0.011	0.032	0.013	-0.030	0.000	-0.018	-0.013
	[0.016]	[0.025]	[0.032]	[0.022]	[0.025]	[0.022]	[0.023]
Second term	0.134***	0.229***	0.315***	0.168***	0.181***	0.175***	0.176***
	[0.016]	[0.028]	[0.039]	[0.022]	[0.025]	[0.022]	[0.022]
TermGPA$_{t-1}$		0.346***	-0.161***	-0.187***	0.094***		0.199***
		[0.021]	[0.014]	[0.038]	[0.035]		[0.072]
LiteratureGPA$_{t-1}$		0.314***				0.023	-0.015
		[0.023]				[0.036]	[0.033]
Female	0.609***	0.220***	0.062***				
	[0.063]	[0.031]	[0.018]				
Ex.absence		-0.096***	-0.065***	-0.054***	-0.087***	-0.058***	-0.065***
		[0.013]	[0.012]	[0.018]	[0.020]	[0.019]	[0.019]
Unex. absence		-0.070***	-0.088***	-0.097***	-0.095***	-0.104***	-0.098***
		[0.024]	[0.023]	[0.025]	[0.026]	[0.026]	[0.026]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	4,918	3,920	3,920	2,928	3,920	2,928	2,928
R-squared	0.088	0.546	0.077	0.088	0.735		

Table A5: *Full sample (2008-2016): Panel data value-added, impact on foreign language GPA and sports GPA from afternoon classes. Columns (1)-(2) use OLS, where (2) adds controls and an autoregressive term. Columns (3)-(4) uses first-difference specification where (3) has only contemporaneous inputs (levels) and (4) adds differencing of time-varying inputs between t and $t-1$, note the reported coefficients in (4) are for the differenced covariates. Column (5) presents the results from fixed effects within estimation and Columns (6)-(7) estimate Arellano-Bond dynamic panel with difference GMM, consequently adding also the previous term average GPA as a pre-determined covariate. Number of students in (6)-(7) are 681 for foreign language GPA and 689 for sports GPA. Instruments for the differenced equation GMM-type: second lag of ForeignGPA or SportsGPA, respectively and first lag of $TermGPA_{t-1}$. All grades and absences are standardised with mean zero and variance one. Fixed effects include separate time-varying indicators for grade and year and time-invariant indicators for profile and teacher. Robust standard errors in brackets, clustered by student. Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
Foreign Language GPA	lanGPA	lanGPA	ΔlanGPA	ΔlanGPA	lanGPA	lanGPA	lanGPA
Afternoon	0.009	0.030	0.022	0.010	-0.000	0.034*	0.028
	[0.016]	[0.024]	[0.028]	[0.020]	[0.022]	[0.021]	[0.021]
Second term	0.061***	0.173***	0.166***	0.113***	0.121***	0.133***	0.139***
	[0.015]	[0.028]	[0.035]	[0.019]	[0.022]	[0.022]	[0.022]
TermGPA_{t-1}		0.333***	-0.165***	-0.229***	0.018		0.022
		[0.024]	[0.014]	[0.034]	[0.036]		[0.062]
ForeignGPA_{t-1}		0.413***				0.094**	0.110***
		[0.024]				[0.041]	[0.039]
Female	0.488***	0.097***	0.083***				
	[0.066]	[0.026]	[0.017]				
Ex. absence		-0.073***	-0.056***	-0.059***	-0.052***	-0.066***	-0.064***
		[0.012]	[0.013]	[0.017]	[0.017]	[0.019]	[0.020]
Unex. absence		-0.109***	-0.138***	-0.086***	-0.087***	-0.103***	-0.094***
		[0.023]	[0.025]	[0.023]	[0.027]	[0.025]	[0.026]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	4,512	3,576	3,576	2,780	3,710	2,646	2,646
R-squared	0.054	0.614	0.076	0.063	0.785		

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. variable:	OLS	OLS	FD	FD	FE	GMM	GMM
Sports GPA	spoGPA	spoGPA	ΔspoGPA	ΔspoGPA	spoGPA	spoGPA	spoGPA
Afternoon	0.028*	0.069***	0.090***	0.029	0.028	0.031	0.029
	[0.015]	[0.022]	[0.029]	[0.020]	[0.020]	[0.020]	[0.020]
Second term	0.128***	0.188***	0.257***	0.140***	0.136***	0.134***	0.131***
	[0.016]	[0.028]	[0.037]	[0.019]	[0.023]	[0.020]	[0.019]
TermGPA_{t-1}		0.060***	-0.082***	-0.114***	0.066		-0.049
		[0.018]	[0.014]	[0.036]	[0.043]		[0.074]
SportsGPA_{t-1}		0.502***				-0.035	-0.048
		[0.031]				[0.054]	[0.050]
Female	-0.128***	-0.087***	0.029*				
	[0.045]	[0.027]	[0.017]				
Ex. absence		-0.025*	-0.018	-0.011	-0.045	-0.023	-0.022
		[0.015]	[0.013]	[0.015]	[0.028]	[0.015]	[0.015]
Unex. absence		-0.019	-0.056**	0.030	0.012	0.030	0.030
		[0.021]	[0.023]	[0.030]	[0.027]	[0.029]	[0.028]
Grade & year FE		x	x	x	x	x	x
Profile & teacher FE		x	x				
Observations	4,643	3,669	3,669	2,731	3,700	2,700	2,700
R-squared	0.011	0.407	0.064	0.057	0.683		

Table A6: *Full sample (2008-2016): Panel data value-added, impact on the math GPA gap (term - year) using within year variation. Columns (1)-(2) use OLS with (2) adding covariates including controls for grade and year, profile and teachers. Columns (3) uses first-difference specification with differencing of time-varying school inputs: note the reported coefficients in (3) are for the differenced covariates. Column (4) presents the results from fixed effects within estimation. Column (5) estimates Arellano-Bond dynamic panel with system GMM. Instruments for the differenced equation GMM-type: second lag of MathGPA respectively and first lag of TermGPA$_{t-1}$. Columns (6)-(8) introduce more flexible syntax allowing for serial correlation in the composite error term. All grades and absences are standardised with mean zero and variance one. Fixed effects include separate time-varying indicators for grade and year and time-invariant indicators for profile and teacher. Robust standard errors in brackets, clustered by student. Columns (6)-(8) use standard errors adjusted for clustering on the student level using the two-step estimator as described in Windmeijer (2005). Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	FD	FE	GMM AB/BB	GMM LDPD	GMM LDPD	GMM LDPD
Dep. variable: **Math GPA gap**	mathGPAgap	mathGPAgap	ΔmathGPAgap	mathGPAgap	mathGPAgap	mathGPAgap	mathGPAgap	mathGPAgap
Afternoon	-0.033***	-0.079***	-0.073***	-0.089***	-0.050***	-0.121***	-0.128***	-0.151***
	[0.008]	[0.015]	[0.017]	[0.017]	[0.014]	[0.022]	[0.023]	[0.024]
Secondterm	0.011	0.072***	0.041***	0.077***	0.088***	0.046***	0.062***	0.061***
	[0.009]	[0.011]	[0.010]	[0.012]	[0.007]	[0.013]	[0.013]	[0.013]
Ex. absence		-0.022***	-0.040***	-0.036***	-0.047***	-0.038***	-0.044***	-0.038***
		[0.003]	[0.010]	[0.007]	[0.007]	[0.010]	[0.011]	[0.011]
Unex. absence		-0.019*	-0.026*	-0.022	-0.023*	-0.036	-0.030	-0.019
		[0.010]	[0.014]	[0.014]	[0.013]	[0.034]	[0.034]	[0.021]
TermGPA$_{t-1}$		-0.030***	-0.257***	-0.121***	-0.016	0.150***	0.203***	0.229***
		[0.004]	[0.019]	[0.012]	[0.016]	[0.042]	[0.063]	[0.070]
MathGPAgap$_{t-1}$					-0.479***	-0.178***	-0.258***	-0.156***
					[0.014]	[0.047]	[0.052]	[0.052]
dgmmiv						L(3/.)	L(5/.)	L(8/.)
lgmmiv						L(2/.)	L(2/.)	L(2/.)
Grade & year FE		x	x	x	x	x	x	x
Profile & teacher FE		x						
Observations	10,860	9,215	7,601	9,215	9,214	6,286	6,286	6,286
R-squared	0.002	0.030	0.077	0.088				
# students					1,600	1,325	1,325	1,325

Table A7: *Full sample (2008-2016): Panel data value-added, impact on the history GPA gap (term - year) using within year variation. Columns (1)-(2) use OLS with (2) adding covariates including controls for grade and year, profile and teachers. Columns (3) uses first-difference specification with differencing of time-varying school inputs: note the reported coefficients in (3) are for the differenced covariates. Column (4) presents the results from fixed effects within estimation. Column (5) estimates Arellano-Bond dynamic panel with system GMM. Instruments for the differenced equation GMM-type: second lag of MathGPA respectively and first lag of TermGPA$_{t-1}$. Columns (6)-(8) introduce more flexible syntax allowing for serial correlation in the composite error term. All grades and absences are standardised with mean zero and variance one. Fixed effects include separate time-varying indicators for grade and year and time-invariant indicators for profile and teacher. Robust standard errors in brackets, clustered by student. Columns (6)-(8) use standard errors adjusted for clustering on the student level using the two-step estimator as described in Windmeijer (2005). Statistical significance: *** p<0.01, ** p<0.05, * p<0.1.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	FD	FE	GMM	GMM	GMM	GMM
					AB/BB	LDPD	LDPD	LDPD
Dep. variable:								
History GPA gap	histGPAgap	histGPAgap	ΔhistGPAgap	histGPAgap	histGPAgap	histGPAgap	histGPAgap	histGPAgap
Afternoon	0.066***	0.119***	0.105***	0.119***	0.049**	0.069**	0.065**	0.048*
	[0.014]	[0.023]	[0.026]	[0.026]	[0.023]	[0.029]	[0.026]	[0.029]
Second term	0.153***	0.198***	0.178***	0.201***	0.144***	0.147***	0.135***	0.170***
	[0.014]	[0.015]	[0.014]	[0.017]	[0.014]	[0.019]	[0.017]	[0.020]
Ex. absence		-0.013***	-0.015	-0.020*	-0.031***	-0.018	-0.025**	-0.030**
		[0.005]	[0.013]	[0.011]	[0.011]	[0.014]	[0.013]	[0.014]
Unex. absence		-0.013*	-0.010	-0.013	-0.002	-0.012	-0.003	-0.004
		[0.007]	[0.008]	[0.009]	[0.005]	[0.014]	[0.009]	[0.007]
TermGPA$_{t-1}$		-0.025***	-0.317***	-0.156***	-0.189***	-0.086***	-0.159***	-0.199***
		[0.004]	[0.029]	[0.021]	[0.021]	[0.031]	[0.026]	[0.034]
HistGPA gap$_{t-1}$					-0.397***	-0.071***	-0.257***	-0.138**
					[0.015]	[0.028]	[0.044]	[0.059]
dgmmiv						L(3/.)	L(5/.)	L(8/.)
lgmmiv						L(2/.)	L(2/.)	L(2/.)
Grade & year FE		x	x	x	x	x	x	x
Profile & teacher FE		x						
Observations	6,963	6,499	5,153	6,499	5,595	5,595	5,595	5,595
R-squared	0.036	0.064	0.116	0.095				
# students					1,336	1,336	1,336	1,336