Economic Inquiry

GENDER PERFORMANCE GAPS: QUASI-EXPERIMENTAL EVIDENCE ON THE ROLE OF GENDER DIFFERENCES IN SLEEP CYCLES

LESTER LUSHER and VASIL YASENOV*

Sleep studies suggest that girls go to sleep earlier, are more active in the morning, and cope with sleep deprivation better than boys. We provide the first causal evidence on how gender differences in sleep cycles can help explain the gender performance gap. We exploit over 240,000 assignment-level grades from a quasi-experiment where students' schedules alternated between morning and afternoon start times each month. Relative to girls, we find that boys' achievement benefits from a later start time. For classes taught at the beginning of the school day, our estimates explain up to 16% of the gender performance gap. (*JEL* H52, I20, I21)

I. INTRODUCTION

Motivated by interests in fairness and equal opportunity, researchers and policymakers frequently discuss the causes of gender differences in educational outcomes. As early as the third grade, performance gaps begin to arise, where boys record lower reading scores than girls. These gaps continue to permeate and grow through secondary school where, for instance, by the age of 17, boys underperform girls in reading by 0.3 standard deviations. This gap is nearly half of the corresponding Black-White performance gap, and is equivalent to approximately 1.5 years of schooling (Dee 2007; Riordan 1999). Girls are also increasingly more likely to attend and graduate from postsecondary schools, with nearly 60% of college students in the United States being female (Diprete and Buchmann 2013; Vedder 2015).

Many studies have sought to identify the mechanisms through which gender performance gaps arise. Early studies primarily focused on the roles of biological differences. While tests of general intelligence suggest no distinctions between boys and girls, there are gender differences on particular cognitive tasks. For instance, boys do better on visual–spatial tasks, while females excel at certain verbal tasks (Neisser et al. 1996). Gender differences in brain structures and in exposure to sex hormones could also influence gender-specific skills (Cahill 2005; Halpern 2013; Kimura 1999; Lippa 2005). More recent studies have investigated the importance of "environmental" factors. For example, boys respond more positively to competitive test-taking environments, whereas women tend to outperform men in less competitive or non-competitive contexts (Niederle and Vesterlund 2010; Ors, Palomino, and Peyrache 2013). Other studies document how students perform better when taking a class with a teacher of the same sex, and hence gender gaps arise when the gender composition of the teachers is unbalanced (Bettinger and Long 2005; Carrell, Page, and West 2010; Dee 2005, 2007; Hoffmann and Oreopoulos 2009; Holmlund and Sund 2008).

Another potential explanation that remains unexplored by social scientists centers on gender differences in sleep cycles. Sleep studies suggest that men have longer circadian periods, or "body clocks," predisposing them to later bedtimes and morning wake-up times. Consequently, women show a stronger inclination for activity earlier in the day than men (e.g., Adan and Natale 2003;

*We would like to thank Scott Carrell, Joseph Doyle, Paco Martorell, Doug Miller, Giovanni Peri, Vladimir Tyazhelnikov, three anonymous referees, and participants at UC Davis brownbag series and the 2016 WEAI Dissertation Workshop for helpful comments. Nicholas Halliwell, Bhaverpreet Sidhu, and Catherine Jiang provided excellent research assistance.

Lusher: Assistant Professor, Department of Economics, University of Hawaii at Manoa, Honolulu, HI 96822. Phone 510 938-5223, Fax 808 956-4347, E-mail lrlusher@ucdavis.edu

Yasenov: Postdoctoral Scholar, Goldman School of Public Policy, University of California, Berkeley, CA 94720; Institute for the Study of Labor (IZA), 53113 Bonn, Germany. Phone 408 510-2794, Fax 510 643-9657, E-mail yasenov@berkeley.edu

ABBREVIATION

STEM: Science, Technology, Engineering, and Mathematics

Economic Inquiry
(ISSN 0095-2583)
Vol. 56, No. 1, January 2018, 252–262

doi:10.1111/ecin.12483
Online Early publication July 26, 2017

Lee, McEnany, and Weekes 1999; Tsai and Li 2004). Furthermore, women on average cope with sleep deprivation better than men, and can rebound more quickly from mild sleep deprivation. Studies also suggest that when sleep deprived, women tend to "catch up" on their sleep better (see Petersen 2011, and articles therein). For instance, in one study, subjects were asked to sleep 6 hours per night over six nights, and then were given two nights of extended overnight sleep. Women slept more on those two nights, and also scored higher on varying post-measures of performance (Breus 2012). Studies also suggest that women spend more time than men in deep, slow-wave sleep stages over the course of a night, and deep sleep is restorative and memory boosting (e.g., Ehlers and Kupfer 1997).[1] The findings from these studies are especially important given the widespread sentiment that students, irrespective of gender, are not getting enough sleep before school.[2] There is ample evidence that sleep and academic achievement are positively correlated (e.g., Pagel, Forister, and Kwiatkowki 2007), and studies from economics have found increases in overall student performance in response to delayed school start times (Carrell, Maghakian, and West 2011; Edwards 2012). If boys receive less sleep than girls and are particularly harmed by lack of sleep, then early school start times could help explain settings where girls outperform boys.

To our knowledge, this paper provides the first causal evidence on the role of gender differences in sleep cycles explaining the gender performance gap. Our data consist of over 240,000 assignment-level grades from a community of middle and high schoolers in an Eastern European country. Exogenous variation comes from a 6-year quasi-experiment where students, by cohort, alternated between morning and afternoon school start times every month. All other aspects were kept constant, including the teachers who taught the classes and the ordering of classes. Students who attended in the morning started at 7:30 a.m., while afternoon start students began at 1:30 p.m. This setting allows us to estimate models with multiple dimensions of fixed effects. Importantly, we can include class fixed effects to control for any unobserved class-level characteristics.[3] Class fixed effects implicitly control for teacher fixed effects since each class is taught by exactly one teacher. They also avoid the need to rely on settings with standardized grading or testing procedures since students within a class complete the same assignments. Furthermore, since start times varied across students within each month, we can include month fixed effects to control for any unobserved factors that vary by month and influence student performance (e.g., weather). Thus, our identification strategy effectively compares the performance of students in the same class across early and late start times while controlling for differential month effects.

Consistent with the hypothesis of gender differences in sleep cycles, we find evidence that boys enjoy a boost in performance relative to girls in response to the later school start time. The estimates for average effects are precisely estimated, ranging between 0.021 and 0.025 standard deviation increases, and are equivalent to increasing a student's teacher quality by roughly a quarter of a standard deviation (Rockoff 2004). Using detailed class schedules, we also test for how the gender differential effect varies by the period of the class. We observe that the relative gain boys receive from the late start are largest in classes taught in the beginning of the school day (7:30–10:00 a.m. in early start months, 1:30–4:00 p.m. in late start months). For these classes, the gender differential response to the late start explains up to 16% of the observed gender performance gap in our setting. The results are consistent across alternative specifications, including ordered logit and ordered probit models.

The remainder of this paper proceeds as follows. Section II introduces the data and the institutional background. Section III discusses our identification strategies and econometric specifications. Section IV presents the results and Section V concludes.

II. DATA AND QUASI-EXPERIMENTAL SETTING

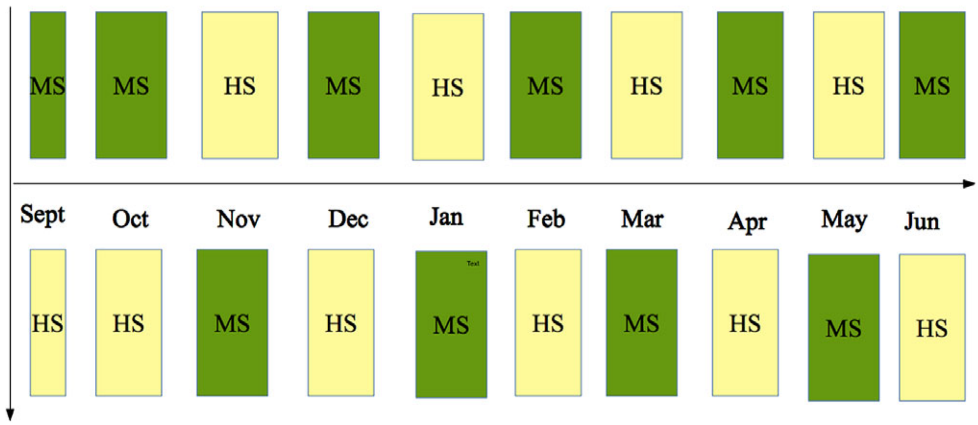
Our study focuses on a community of middle and high schoolers from 2008 until 2014. Each

1. One limitation to these physiological studies is the arguably narrow generalizability of the results, where individuals who select into these studies are likely different from the student population of interest. For instance, to our knowledge, all prior physiological studies were conducted on consenting adults.

2. Read Kelley and Lee (2014) for a recent summary of sleep research and policy discussion as related to education.

3. We define "class" as a combination of course (e.g., tenth-grade Biology), school year (e.g., 2009–2010), and lecture room. Each lecture room is occupied by only one cohort of students, while all incoming middle and high school students are assigned to one of several possible cohorts.

FIGURE 1
Quasi-Experimental Setting



HS, high school; MS, middle school.

incoming middle and high schooler gets assigned a cohort, and students only take classes with other students from their cohort for the entirety of their time in school. Students do not have the ability to switch cohorts or select into courses.[4] The data comprise a complete list of raw, pen-to-paper grades received on all homework, quiz, and exam assignments. Each assignment received one of five possible integer grades, ranging from 2 (*lowest*) to 6 (*highest*). Raw grades were not curved or edited upon being graded. For our primary analyses, we normalize grades to a mean of zero and a standard deviation of one within class.

The school day in our setting is highly structured and very simple. Each day consists of seven 40-minute periods. A 10-minute break is given between each period except for between periods 3 and 4, where instead a 20-minute recess is given. Each class typically lasts the length of one period, though some classes cover two periods, with the corresponding break still intact. The period at which a class is taught may vary by weekday and by semester. Students are sometimes given an "off-period" during period 7 such that they have no class to attend and can leave school early.

During the period of our study, students, by cohort, alternated between morning and afternoon school start times each month. All other aspects of the schools were kept constant, including the period of the classes, the locations of the classroom, and the teachers who taught the classes. High school cohorts started at 7:30 a.m. during September and the "even" months (October, December, February, April, and June), while middle school cohorts attended at 7:30 a.m. in all remaining "odd-numbered" months (November, January, March, and May). Thus, high (middle) school cohorts started school in the afternoon during "odd" (September and "even") months. The afternoon block started at 1:30 p.m. (see Figure 1).[5] The quasi-experiment was implemented in response to facility constraints and local organizers' inabilities to come to an agreement where cohorts remained entrenched in one block for the entire school year.[6]

Summary statistics are presented in Table 1. Nearly 45% of the students are male. The average class size was over 23 students. Girls received an average grade of 4.47, while boys received a 4.13. This gender performance gaps is reduced

4. Students are assigned to their cohort based on their stated educational interests upon entering the middle or high school. Students stated interests are mapped into three broad categories, each of which defines the "specialization" of the cohort: languages, arts, or math. Although cohorts differ in specialization, over 80% of subjects that students take overlap across all cohorts, or in other words, less than one in every five subjects students take are unique to that student's cohort. This is typical for the majority of public schools in the country we study. The country of origin is a member of the European Union. The average graduating class size is approximately 120 students.

5. The first day of school is typically in the middle of September.

6. Lusher and Yasenov (2016) utilize the same setting to investigate student performance in double-shift schooling systems.

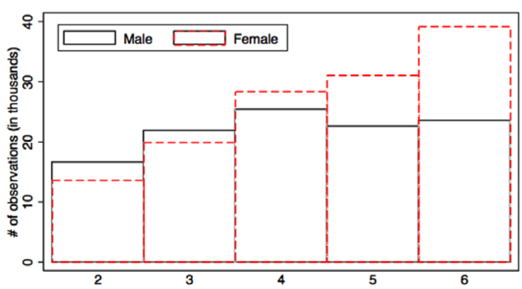
TABLE 1
Summary Statistics

	Mean	SD
Student level [$N = 1{,}110$]		
Male	0.444	(0.497)
Native ethnicity	0.778	(0.416)
Class level [$N = 1{,}086$]		
# of students	23.412	(3.470)
Male teacher	0.138	(0.345)

Assignment-level grades	**Female Students**	**Male Students**	**Performance Gap**
Full sample	4.472	4.132	0.340
[$N = 241{,}945$]	(1.327)	(1.358)	(0.005)
STEM courses	4.066	3.780	0.286
[$N = 79{,}701$]	(1.353)	(1.344)	(0.010)
Classes w/male teacher	4.449	4.157	0.291
[$N = 26{,}204$]	(1.367)	(1.361)	(0.017)
Late start months	4.444	4.113	0.329
[$N = 128{,}880$]	(1.326)	(1.356)	(0.007)

Notes: Each cell under "Female Students" and "Male Students" reports the mean assignment grade with standard deviations presented below in parentheses. Under "Performance Gap," each cell reports the difference in mean assignment grade between girls and boys with the standard error for the mean displayed in parentheses. Assignment grades could take on one of five integer values between 2 and 6.

FIGURE 2
Histogram of Assignment Grades



when focusing solely on assignments that were completed during late start months, as well as on classes in a science, technology, engineering, and mathematics (STEM) field or which were taught by a male teacher. Figure 2 displays the distribution of assignment scores by gender, and shows that girls had a higher probability than boys of attaining the highest possible assignment scores of 5 and 6, while boys were more likely to receive a 2, 3, or 4.

III. IDENTIFICATION STRATEGY

We first consider analysis using the standardized assignment grades. By normalizing to a mean of zero and a standard deviation of one within class and utilizing within class, across assignment variation, we effectively account for across class differences in difficulty or grading standards. This also allows our estimated treatment effects to be comparable to those from other studies utilizing standardized outcome variables. As a robustness check, in Section IV.C, we focus on the raw assignment grades, each consisting of five possible discrete values (2, 3, 4, 5, and 6), to fit ordered probit and ordered logit models.

Our main analysis estimates the following specification:

$$(1) \quad Grade_{aicmy} = \beta \left(Male_i \times LateStart_{im} \right) + \mathbf{x}'_{aicmy}\gamma + \delta_{cy} + \lambda_m + \epsilon_{aicmy}$$

where $Grade_{aicmy}$ is the normalized grade student i received on assignment a in course c during month m and school year y. $Male_i$ is an indicator for whether student i was male. $LateStart_{im}$ is an indicator variable equal to one if student i's assignment was completed during a late start month. $\mathbf{x}_{aicmy}$ is a vector of controls, including $Male_i$, $LateStart_{im}$, an indicator for student i's race, and the order of the assignment a and the number of assignments student i completed in class cy within month m. We also include an interaction term between $Male_i$ and an indicator for class cy's teacher being male, which can be interpreted as the student–teacher gender interaction effect (Bettinger and Long 2005; Carrell, Page, and West 2010; Dee 2005, 2007; Hoffmann and Oreopoulos 2009; Holmlund and Sund 2008). This lets us compare our estimated gender interaction effects with prior work, and to juxtapose the magnitudes of our estimates for β with another major determinant of gender performance gaps.

The quasi-experimental setting grants us the opportunity to include multiple dimensions of fixed effects, all of which were not permeable in the previous relevant literature, and which effectively eliminate any potential concerns for endogeneity bias. The core of our identification strategy centers on class fixed effects δ_{cy}, which control for unobserved factors that vary at the class level and affect student performance. Importantly, they absorb teacher fixed effects since each class is taught by exactly one teacher across all months. Class fixed effects also avoid the need to rely on settings with standardized

TABLE 2
Main Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male	−0.217***	−0.236***	−0.236***	−0.237***	−0.253***	−0.252***	−0.258***	−0.258***
	(0.011)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	(0.013)
Male X								
Late start		0.025***	0.023***	0.022***	0.025***	0.023***	0.021***	0.021***
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Male X								
Male teacher		0.053	0.052	0.045	0.057	0.058*	0.050	0.051
		(0.033)	(0.033)	(0.033)	(0.035)	(0.035)	(0.036)	(0.036)
Late start		−0.053***	−0.051***	−0.053***	−0.053***	−0.050***	−0.049***	−0.053***
		(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Observations	241,945	241,945	241,945	241,945	241,945	241,945	241,945	241,945
R^2	0.012	0.012	0.017	0.031	0.013	0.018	0.029	0.034
Class FE					X	X	X	X
Month FE			X	X		X		X
Controls				X			X	X

Notes: Each column within each panel pertains to a single regression, while each row corresponds to a regressor of interest. Standard errors are clustered at the class level and presented in parentheses. FE, fixed effects.
$^*p<.10$, $^{**}p<.05$, $^{***}p<.01$.

grading or testing procedures across classes since students within a class are completing the exact same assignments and tests. Thus, we are solely comparing the academic performances of boys and girls within the same class and subjecting the students to the same class-level shocks, such as the teacher's characteristics (e.g., ability/experience) or class size/difficulty. Furthermore, since start time varies across students within month, we can include month fixed effects λ_m to control for any unobserved variables that vary by month and influence student performance (e.g., weather).

The coefficient β can be interpreted as the relative benefit males gain over females by having a later school start time. In total, in our identification of β, we compare the performance of students in the same class across early and late start times while controlling for differential month effects. Our estimates for β will be biased only if an omitted term correlates across every odd month, has power in predicting assignment grades, and differentially impacts high school versus middle school cohorts.

IV. RESULTS

A. Main Results

Our main results are reported in Table 2. Column 1 presents a simple, bivariate regression of standardized grade on gender to show that unconditionally, girls earn on average an assignment grade 0.22 standard deviations higher than boys. The remaining columns include the interaction term $Male_i \times LateStart_{im}$ and an interaction term between $Male_i$ and an indicator for the class teacher being male. Across columns, we check the sensitivity of the results to the inclusion of class fixed effects δ_{cy}, month fixed effects λ_m, and controls $\mathbf{x}_{aicmy}$.

Consistent with the sleep cycle literature, we find that boys enjoy a boost in performance relative to girls in response to the later school start time. Across specifications, we attain statistical significance at the 1% level for all estimates of β. The magnitudes of the estimates are fairly consistent across specifications, culminating in a 0.021 standard deviation effect size under our fully specified model. Although less precisely estimated, we also find student–teacher gender interaction effects between 0.045 and 0.058 standard deviations across specifications, effect sizes which are mostly in line with the majority of previous studies.[7]

Given the uniqueness of our setting, it is useful to juxtapose the magnitude of our results against other prominent determinants of student achievement. Across specifications within our own study, the gender differential response to the late start is, on average, nearly half the magnitude of the

7. For instance, Dee (2007) and Hoffmann and Oreopoulos (2009) find a gender interaction effect of 0.054 and 0.50 standard deviations, respectively. The majority of gender interaction studies also center on specifications with class fixed effects.

student–teacher gender interaction effect. Estimating the effects of morning versus afternoon classes holding start time fixed, Pope (2015) found a very similar effect size of 0.024 standard deviations on standardized math test scores.[8] Our estimated effect sizes are equivalent to increasing a student's teacher quality by roughly a quarter of a standard deviation (Rockoff 2004). Using estimates from the peer effects literature (Feld and Zolitz 2017), our estimated effect sizes are comparable to increasing the average ability of a student's classmates by 1.67 standard deviations.

B. Results by Subsamples

Table 3 estimates our primary specification on different subsamples of interest in order to investigate potential heterogeneities. The top panel "Early Classes" focuses on the subset of classes which took place during periods 1, 2, or 3 across all weekdays. That is, we focus solely on assignments completed in classes that were consistently taught between 7:30 and 10:00 a.m. during early start months and between 1:30 and 4:00 p.m. during late start months across all weekdays. These are classes for which a later start could lead to an especially relieving effect since they are taught at times during early start months for which students likely prefer sleeping. The estimates for gender differences in response to the late start increase by over 50% across specifications, ranging between 0.032 to 0.038 standard deviations. Furthermore, when focusing on the subsample of classes that were taught during periods 4, 5, 6, or 7 across all weekdays ("Late Classes"), the estimates become statistically indistinguishable from zero while maintaining relatively tight confidence intervals. We interpret these results as further evidence that the relatively poor male performance is at least partially due to a lack of sleep, and poor response to a lack of sleep, when taking early morning classes. Across specifications, the effect size for the differential gender response to the late start ranges between 12% and 16% of the observed gender performance gap.

Still, an important limitation of our study is that the effect we find could potentially be attributable to factors other than gender differential responses to sleep. For one, we do not observe actual sleep behavior. Moreover, potential afternoon-class effects could exist, where holding school start time constant, it could be that student performance differs in the morning versus afternoon. Then, our effects would be explained by a gender differential response to an afternoon-class effect instead of a gender differential response to a later start time. A study from Pope (2015) investigates this possibility. First, Pope (2015) finds that holding start time constant, students overall perform better in the morning (roughly 8:00 to 10:00 a.m.) versus the afternoon (roughly 12:50 to 2:45 p.m.). Then, the author presents some suggestive evidence that this effect differs by gender, where the effect for two of his four academic achievement outcome variables is statistically significantly different across genders at the 10% level. However, it turns out that the positive morning-class effect is *stronger* for boys versus girls, or conversely, the "drop" in performance in the afternoon relative to the morning is smaller for girls versus boys. This suggests that if anything our results could be underestimating the relative benefit males gain from a later school start time since our coefficient could capture both a late start effect and an afternoon-class effect.[9]

The bottom two panels of Table 3 consider STEM and non-STEM course subsamples. First, as is well documented in other studies, we find that boys receive a significant boost in relative performance when focusing on STEM courses (e.g., Carrell, Page, and West 2010). Second, we see that the gender differential response to the late start is primarily driven by non-STEM courses. Finally, the student–teacher gender interaction effects are significantly larger in non-STEM versus STEM courses. Results from previous studies investigating student–teacher gender interaction effects have been mixed when considering different courses.[10]

Finally, a natural heterogeneity to investigate is by age. Since sleep differences are tied to the hormonal cycle, which in turn goes through major changes during puberty, one might expect to see varying gender differential responses for

8. Pope (2015) found no effect (0.003 standard deviations) on standardized English test scores for morning verses afternoon classes, conditional on school start time.

9. Table 4 also shows the coefficient of the indicator for late start. Discussed in more detail in Lusher and Yasenov (2016), we find that similar to Pope (2015), students overall perform slightly better in morning classes.

10. For instance, in response to having a female teacher, Dee (2007) finds girls do better in history, but worse in math, with no differential effect in science or English, while boys do worse with female teachers in math and science courses. On the other hand, Carrell, Page, and West (2010) find that females respond positively to female teachers in STEM courses.

TABLE 3
Main Results by Subsamples

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Early classes								
Male	−0.217***	−0.245***	−0.243***	−0.247***	−0.262***	−0.261***	−0.267***	−0.267***
	(0.021)	(0.025)	(0.025)	(0.025)	(0.027)	(0.027)	(0.027)	(0.027)
Male X								
Late start		0.038**	0.035**	0.033*	0.037**	0.035**	0.033*	0.032*
		(0.018)	(0.017)	(0.017)	(0.018)	(0.018)	(0.017)	(0.017)
Male X								
Male teacher		0.063	0.062	0.059	0.070	0.070	0.065	0.064
		(0.047)	(0.048)	(0.047)	(0.051)	(0.051)	(0.048)	(0.049)
Observations	47,146	47,146	47,146	47,146	47,146	47,146	47,146	47,146
R^2	0.012	0.013	0.019	0.035	0.014	0.020	0.034	0.039
Late classes								
Male	−0.211***	−0.236***	−0.237***	−0.236***	−0.255***	−0.256***	−0.258***	−0.259***
	(0.020)	(0.023)	(0.023)	(0.023)	(0.025)	(0.025)	(0.024)	(0.025)
Male X								
Late start		0.015	0.016	0.015	0.015	0.016	0.013	0.015
		(0.018)	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Male X								
Male teacher		0.105**	0.104**	0.093*	0.116**	0.119**	0.104*	0.107*
		(0.051)	(0.051)	(0.053)	(0.054)	(0.055)	(0.057)	(0.057)
Observations	55,949	55,949	55,949	55,949	55,949	55,949	55,949	55,949
R^2	0.011	0.012	0.019	0.030	0.013	0.020	0.025	0.032
STEM courses								
Male	−0.165***	−0.166***	−0.166***	−0.171***	−0.178***	−0.178***	−0.186***	−0.186***
	(0.018)	(0.019)	(0.019)	(0.019)	(0.021)	(0.021)	(0.020)	(0.020)
Male X								
Late start		0.008	0.007	0.008	0.008	0.007	0.007	0.008
		(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Male X								
Male teacher		−0.056	−0.056	−0.072	−0.054	−0.054	−0.069	−0.069
		(0.076)	(0.076)	(0.079)	(0.079)	(0.079)	(0.083)	(0.083)
Observations	79,701	79,701	79,701	79,701	79,701	79,701	79,701	79,701
R^2	0.007	0.008	0.014	0.027	0.008	0.015	0.023	0.029
Non-STEM courses								
Male	−0.242***	−0.273***	−0.273***	−0.273***	−0.293***	−0.292***	−0.297***	−0.297***
	(0.014)	(0.017)	(0.017)	(0.016)	(0.018)	(0.018)	(0.017)	(0.017)
Male X								
Late start		0.034***	0.033***	0.030***	0.034***	0.032***	0.029***	0.029***
		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Male X								
Male teacher		0.096***	0.096***	0.091**	0.103***	0.105***	0.097**	0.099**
		(0.037)	(0.037)	(0.037)	(0.040)	(0.040)	(0.040)	(0.040)
Observations	162,244	162,244	162,244	162,244	162,244	162,244	162,244	162,244
R^2	0.015	0.015	0.020	0.034	0.017	0.021	0.033	0.038
Class FE					X	X	X	X
Month FE			X	X		X		X
Controls				X			X	X

Notes: Each panel considers a subsample of interest. Each column within a panel pertains to a single regression, while each row corresponds to a regressor of interest. "Early Classes" were classes taught during periods 1–3 across all weekdays. "Late Classes" were classes taught during periods 4–7 across all weekdays. "STEM courses" includes mathematics, engineering, and science courses, while "Non-STEM courses" focus on all remaining courses. Standard errors are clustered by class and presented in parentheses. FE, fixed effects.

* $p < .10$, ** $p < .05$, *** $p < .01$.

TABLE 4
Main Results by Middle versus High Schoolers

	(1)	(2)	(3)	(4)	(5)
Middle schoolers					
Male	−0.308***	−0.314***	−0.308***	−0.331***	−0.324***
	(0.018)	(0.019)	(0.018)	(0.020)	(0.019)
Male X					
Late start		0.016	0.014	0.016	0.014
		(0.013)	(0.013)	(0.013)	(0.013)
Male X					
Male teacher		−0.087	−0.092	−0.070	−0.077
		(0.064)	(0.061)	(0.065)	(0.062)
Observations	74,249	74,249	74,249	74,249	74,249
R^2	0.024	0.024	0.039	0.026	0.041
High schoolers					
Male	−0.177***	−0.182***	−0.188***	−0.198***	−0.210***
	(0.014)	(0.017)	(0.017)	(0.018)	(0.018)
Male X					
Late start		0.006	0.007	0.006	0.007
		(0.010)	(0.010)	(0.010)	(0.010)
Male X					
Male teacher		0.016	0.012	0.019	0.017
		(0.035)	(0.035)	(0.037)	(0.038)
Observations	167,696	167,696	167,696	167,696	167,696
R^2	0.008	0.008	0.022	0.009	0.026
Class FE				X	X
Month FE					
Controls			X		X

Notes: Each panel considers a subsample of interest. Each column within a panel pertains to a single regression, while each row corresponds to a regressor of interest. Note that month fixed effects cannot be estimated with these subsamples since variation in late start within month across assignments occurs primarily across middle and high schoolers. FE, fixed effects.

$^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

middle schoolers versus high schoolers. One limitation with the variation we exploit is that the month fixed effects in Equation (1) cannot be estimated if we split by age. This is because the variation across students and within month is collinear with cohort, where middle versus high school cohorts alternated start times with each other every month. Hence, results by age are arguably more likely to suffer from endogeneity bias. Nevertheless, by age in Table 4, we estimate Equation (1) without month fixed effects for middle versus high school subsamples. First, we find the gender performance gap is significantly larger for middle schoolers in our setting. Meanwhile, the gender differential response to the late start is positive but statistically insignificant for both samples, and only slightly larger for middle schoolers.

C. *Alternative Specifications*

In order to test the robustness of our results, we consider alternative specifications utilizing the raw assignment grade. Each assignment received an integer grade between 2 and 6. Panel A in Table 5 displays the estimated coefficients from ordered logit and ordered probit regressions.[11] We observe that, similar to Table 2, being a male is associated with a negative effect on the (latent) outcome, while being a male during late start months has a statistically significant positive effect relative to being female. Finally, the student–teacher gender interaction effect has a positive, but statistically insignificant, effect on the latent outcome.

Panel B in Table 5 shows the marginal effects of the variables of interest on the probability of obtaining each possible assignment grade. Marginal effects offer greater insight than the coefficient estimates in Panel A, as they provide a more straightforward interpretation. Here, $y=2$ and $y=6$ refer to the lowest and highest possible assignment grade, respectively. The estimates across both specifications are nearly identical and exhibit similar patterns. Male students are more likely to receive lower assignment grades

11. See Cameron and Trivedi (2005) for a more thorough description of the setup and estimation of ordered discrete data models.

TABLE 5
Alternative Specifications Utilizing Assignment Grades

	Panel A: Coefficients	
	Ordered Logit	**Ordered Probit**
Male	−0.477***	−0.283***
	(0.025)	(0.015)
Male X		
Late start	0.026*	0.015*
	(0.015)	(0.009)
Male X		
Male teacher	0.040	0.030
	(0.065)	(0.038)
N	241,945	241,945
log L	−378,073.86	−378,227.27
Pseudo R^2	0.012	0.011

	Panel B: Marginal Effect				
	Pr($y=2$)	**Pr($y=3$)**	**Pr($y=4$)**	**Pr($y=5$)**	**Pr($y=6$)**
Ordered logit					
Male	0.050***	0.048***	0.021***	−0.029***	−0.091***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.005)
Male X					
Late start	−0.003*	−0.003*	−0.001*	0.002*	0.005*
	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)
Male X					
Male teacher	−0.004	−0.004	−0.002	0.002	0.008
	(0.007)	(0.007)	(0.003)	(0.004)	(0.013)
Ordered probit					
Male	0.057***	0.041***	0.016***	−0.022***	−0.091***
	(0.003)	(0.002)	(0.001)	(0.001)	(0.005)
Male X					
Late start	−0.003*	−0.002*	−0.001*	0.002*	0.005*
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)
Male X					
Male teacher	−0.006	−0.004	−0.002	0.002	0.010
	(0.008)	(0.006)	(0.002)	(0.003)	(0.012)

Notes: Panel A shows the estimated coefficients from ordered logit and ordered probit models while Panel B displays the marginal effects of the regressors of interest for each possible assignment grade, evaluated at the controls' means. Controls include student ethnicity, the order of the assignment, and the number of assignments completed within the same month. Standard errors are clustered on the class level.

*$p < .10$, **$p < .05$, ***$p < .01$.

than girls, reflecting the overall performance gap. Boys also record a higher probability of getting a better grade relative to girls in response to the late start. For instance, girls are 9.1% more likely to receive the highest possible grade of a 6 than boys, while boys receive an additional 0.5% boost over girls in the probability of receiving grade of a 6 in response to the afternoon start time. Conversely, boys are 0.3% less likely to receive the lowest grade of a 2 compared with girls in response to the late start. All in all, these results confirm the findings presented in Section IV.A.[12]

V. CONCLUSIONS

Gender performance gaps are prominent in primary and secondary schools, where girls tend to outperform boys. In this paper, we identify a novel mechanism through which gender gaps could develop: gender differences in sleep cycles. An abundance of sleep studies suggest that circadian rhythms predispose boys to sleeping later and waking up later than girls, and that girls cope with sleep deprivation better than boys. Given the widespread sentiment that overall students are not receiving enough sleep, these studies suggest that

12. In addition, we performed two final robustness checks. First, we calculated the average raw grade for boys and girls for each cohort across all assignments within a class and by month-year, and then estimated Equation (1) but without assignment-level controls. Second, we used the rank ordering of performance for each student within each class and month-year as an outcome variable. The results, reported in the Appendix, are consistent with the findings in the main tables: academic performance for boys increases relative to girls in response to the late start.

early school start times could be especially detrimental to boys.

We find increases in the performance of boys relative to girls in response to a later school start time. Average gender differential effects in response to the late start range between 0.021 and 0.025 standard deviations. We also find that the effects are particularly driven by the subsample of classes taught at the beginning of the school day, classes which are most likely to be affected by early start times. For these classes, the gender differential response to the later start explains nearly a sixth of the gender performance gap.

APPENDIX

TABLE A1
Robustness Check—Average Grades by Gender, Class, and Month

	(1)	(2)	(3)	(4)	(5)
Male	−0.287***	−0.310***	−0.310***	−0.306***	−0.306 w
	(0.015)	(0.017)	(0.017)	(0.018)	(0.018)
Male X					
Late start		0.021*	0.022*	0.021*	0.022*
		(0.012)	(0.012)	(0.013)	(0.013)
Male X					
Male teacher		0.086**	0.087**	0.093**	0.094**
		(0.039)	(0.039)	(0.041)	(0.041)
Observations	19,598	19,598	19,598	19,598	19,598
	0.028	0.031	0.048	0.500	0.521
Class FE				X	X
Month FE			X		X
Controls					

Notes: Each column within a panel pertains to a single regression, while each row corresponds to a regressor of interest. The level of observation is gender by class by month. The outcome variable is the average grade for boys and girls within a cohort across all assignments within a class and by month-year. Note that the units of the outcome variable are different than the ones of Tables 2, 3, and 4. FE, fixed effects.

* $p < .10$, ** $p < .05$, *** $p < .01$.

TABLE A2
Robustness Check—Rank Grades by Class and Month

	(1)	(2)	(3)	(4)	(5)
Male	−2.081***	−2.349***	−2.307***	−2.346***	−2.313***
	(0.251)	(0.246)	(0.245)	(0.173)	(0.174)
Male X					
Late start		0.222	0.145	0.162	0.097
		(0.141)	(0.138)	(0.115)	(0.112)
Male X					
Male teacher		1.030**	1.045**	0.874***	0.868***
		(0.458)	(0.457)	(0.334)	(0.332)
Observations	241,945	241,945	241,945	241,945	241,945
R^2	0.004	0.014	0.016	0.470	0.473
Class FE				X	X
Month FE			X		X
Controls					

Notes: Each column pertains to a single regression, while each row corresponds to a regressor of interest. The outcome variable is the rank within a cohort across all assignments within a class. Note that the units of the outcome variable are different than the ones of Tables 2, 3 and 4. FE, fixed effects.

* $p < .10$, ** $p < .05$, *** $p < .01$.

REFERENCES

Adan, A., and V. Natale. "Gender Differences in Morningness–Eveningness Preference." *Chronobiology International*, 19, 2002, 709–20.

Bettinger, E. P., and B. T. Long. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review*, 95(2), 2005, 152–57.

Breus, M. "Men and Women: Different When It Comes to Sleep." February 28, 2012. Accessed March 14, 2016. https://www.psychologytoday.com/blog/sleep-newzzz/201202/men-and-women-different-when-it-comes-sleep

Cahill, L. "His Brain, Her Brain." *Scientific American*, 292, 2005, 40–47.

Cameron, A. C., and P. K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press, 2005.

Carrell, S. E., T. Maghakian, and J. E. West. "A's from Zzzz's? The Causal Effect of School Start Time on the Academic Achievement of Adolescents." *American Economic Journal: Economic Policy*, 3, 2011, 62–81.

Carrell, S. E., M. E. Page, and J. E. West. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics*, 125, 2010, 1101–44.

Dee, T. S. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review*, 95, 2005, 158–65.

———. "Teachers and the Gender Gaps in Student Achievement," *Journal of Human Resources*, 42, 2007, 528–54.

Diprete, T. A., and C. Buchmann. *The Rise of Women: The Growing Gender Gap in Education and What It Means for American Schools*. New York: Russell Sage Foundation, 2013.

Edwards, F. "Early to Rise? The Effect of Daily Start Times on Academic Performance." *Economics of Education Review*, 31, 2012, 970–83.

Ehlers, C., and D. Kupfer. "Slow-Wave Sleep: Do Young Adult Men and Women Age Differently?" *Journal of Sleep Research*, 6, 1997, 211–5.

Feld, J., and U. Zolitz. "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects." *Journal of Labor Economics*, 35(2), 2017, 387–428.

Halpern, D. F. *Sex Differences in Cognitive Abilities*. New York and London: Psychology Press, 2013.

Hoffmann, F., and P. Oreopoulos. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." *Journal of Human Resources*, 44, 2009, 479–94.

Holmlund, H., and K. Sund. "Is the Gender Gap in School Performance Affected by the Sex of the Teacher?" *Labour Economics*, 15, 2008, 37–53.

Kelley, P., and C. Lee. "Later Education Start Times in Adolescence: Time for Change." *Student Health.* Denver, CO: Education Commission of the States, 2014, 1–5.

Kimura, D. *Sex and Cognition*. Cambridge: MIT Press, 1999.

Lee, K. A., G. McEnany, and D. Weekes. "Gender Differences in Sleep Patterns for Early Adolescents." *Journal of Adolescent Health*, 24, 1999, 16–20.

Lippa, R. A. *Gender, Nature, and Nurture*. New York and London: Routledge, 2005.

Lusher, L., and V. Yasenov. "Double-Shift Schooling and Student Success: Quasi-Experimental Evidence from Europe." *Economics Letters*, 139, 2016, 36–9.

Neisser, U., G. Boodoo, T. J. Bouchard Jr., A. W. Boykin, N. Brody, S. J. Ceci, D. F. Halpern, J. C. Loehlin, R. Perloff, R. J. Sternberg, and S. Urbina. "Intelligence: Knowns and Unknowns." *American Psychologist*, 51(2), 1996, 77–101.

Niederle, M., and L. Vesterlund. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives*, 24, 2010, 129–44.

Ors, E., F. Palomino, and E. Peyrache. "Performance Gender Gap: Does Competition Matter?" *Journal of Labor Economics*, 31, 2013, 443–99.

Pagel, J. F., N. Forister, and C. Kwiatkowki. "Adolescent Sleep Disturbance and School Performance: The Confounding Variable of Socioeconomics." *Journal of Clinical Sleep Medicine*, 3, 2007, 19–23.

Petersen, A. "A Sleep Battle of the Sexes." *Wall Street Journal*, August 23, 2011. Accessed March 16, 2016. http://www.wsj.com/articles/SB10001424053111904279004576524321377942288.

Pope, N. G. "How the Time of Day Affects Productivity: Evidence from School Schedules." *Review of Economics and Statistics*, 98, 2015, 1–11.

Riordan, C. "The Silent Gender Gap," *Education Week*, November 17, 1999. Accessed June 22, 2017. http://www.edweek.org/ew/articles/1999/11/17/12riordan.h19.html

Rockoff, J. E. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94, 2004, 247–52.

Tsai, L.-L., and S.-P. Li. "Sleep Patterns in College Students: Gender and Grade Differences." *Journal of Psychosomatic Research*, 56, 2004, 231–37.

Vedder, R. "The Disappearing College Male." *Forbes Online*, May 4, 2015. Accessed March 16, 2016. https://www.forbes.com/sites/ccap/2015/05/04/the-disappearing-college-male/#b89df284d3c1