



OPEN

A 4-year longitudinal study investigating the relationship between flexible school starts and grades

Anna M. Biller^{1,2,6}✉, Carmen Molenda¹, Fabian Obster^{3,7}, Giulia Zerbini^{1,4}, Christian Förtsch⁵, Till Roenneberg^{1,8} & Eva C. Winnebeck^{1,9}✉

The mismatch between teenagers' late sleep phase and early school start times results in acute and chronic sleep reductions. This is not only harmful for learning but may reduce career prospects and widen social inequalities. Delaying school start times has been shown to improve sleep at least short-term but whether this translates to better achievement is unresolved. Here, we studied whether 0.5–1.5 years of exposure to a *flexible* school start system, with the *daily* choice of an 8 AM or 8:50 AM-start, allowed secondary school students ($n = 63$ –157, 14–21 years) to improve their quarterly school grades in a 4-year longitudinal pre-post design. We investigated whether sleep, changes in sleep or frequency of later starts predicted grade improvements. Mixed model regressions with 5111–16,724 official grades as outcomes did not indicate grade improvements in the flexible system per se or with observed sleep variables nor their changes—the covariates academic quarter, discipline and grade level had a greater effect in our sample. Importantly, our finding that intermittent sleep benefits did not translate into detectable grade changes does not preclude improvements in learning and cognition in our sample. However, it highlights that grades are likely suboptimal to evaluate timetabling interventions despite their importance for future success.

During adolescence, teenagers undergo a plethora of biological and socially-driven developments that also influence their sleep–wake behaviour^{1–4}. Their internal phase (chronotype) delays progressively with age until around 21⁵, while sleep pressure (the homeostatic load) likely accumulates more slowly across the day compared to adults^{6,7}. This predisposes teenagers, more than younger children or adults, to delay into late evening hours thereby also delaying their sleep timing. Early school start times cut teenagers' sleep artificially short in the morning, forcing them to get up before they reach healthy amounts of 8–10 h of night-time sleep on schooldays. On weekends, teenagers sleep not only longer but also later which better suits their delayed circadian clock. This sleep timing difference between school and free days is called “social jetlag”, since it is a constant jetlag situation induced by social schedules^{8,9}. These widespread sleep restrictions are not only connected to compromised health and well-being^{10–14}, but also reduced cognition (e.g. constructive¹⁵ and creative thinking skills^{16,17} or verbal fluency^{18,19}) and decreased academic performance²⁰.

There is now substantial amount of evidence that delayed school start times help to increase sleep durations towards more healthy amounts, at least in the short-term^{21–24}. However, it is less clear whether the increased sleep durations and better sleep quality also translate to better learning in the same students. This could reasonably be expected, given the prominent role of sleep for memory, sustained attention and concentration e.g.^{25–27}.

¹Institute of Medical Psychology, Ludwig Maximilian University Munich, Munich, Germany. ²Graduate School of Systemic Neurosciences, Ludwig Maximilian University Munich, Munich, Germany. ³Statistical Consulting Unit, Department of Statistics, Ludwig Maximilian University Munich, Munich, Germany. ⁴Present address: Department of Medical Psychology and Sociology, University of Augsburg, Augsburg, Germany. ⁵Biology Education, Faculty of Biology, Ludwig Maximilian University Munich, Munich, Germany. ⁶Present address: Institute of Psychology, Bundeswehr University Munich, Munich, Germany. ⁷Present address: Department of Business Administration, Bundeswehr University Munich, Munich, Germany. ⁸Present address: Institute and Polyclinic for Occupational-, Social- and Environmental Medicine, Ludwig Maximilian University Munich, Munich, Germany. ⁹Present address: Chair of Neurogenetics, Faculty of Medicine, Technical University of Munich, and Institute of Neurogenetics, Helmholtz Center Munich, Munich, Germany. ✉email: anna.biller@med.uni-muenchen.de; eva.winnebeck@helmholtz-muenchen.de

Thus, several studies previously tried to evaluate the effects of delayed school start times on grades or scores. These studies yielded very mixed results, probably due to differences in study designs, interventions, exposure times and outcome measures²⁸. However, all interventions assessed were static changes in school start times whereas the possibility to make school start times flexible has been largely overlooked although previously raised in a commentary²⁹.

Here, we studied whether a *flexible* school start system, as implemented in a secondary school in Germany^{30,31}, and concurrent changes in sleep were associated with changes in grades in multiple academic disciplines. The flexible system entailed that the school changed from a permanent fixed start at mainly 8 AM to a *flexible* school start that allowed senior students to choose *daily* whether to attend school at 8 AM or skip the first class and start at 8:50 AM. The flexible system was associated with a variety of short- and long-term changes in students' sleep patterns on which we report in detail elsewhere^{30,31}. In summary, on days with later starts, students slept on average about 1 h longer by delaying their offset times by 1 h but maintaining the same average onset times. Subjective sleep quality was improved, and the amount of alarm-driven waking slightly reduced. This was seen both immediately after changing into the flexible system³⁰ as well as after 1 full year in the system³¹. Nonetheless, sleep patterns did not improve from the conventional to the flexible system overall since, on average, students did not accumulate enough late starts during the time of monitoring to result in a significant difference of the weekly average. However, there was wide variation between students on the number of late starts chosen and the sleep gain or loss achieved^{30,31}.

In the present study, we analysed students' quarterly grades from 12 academic school subjects across 4 years to examine effects of this new system on academic grades in detail. With 2.5 years of data prior and up to 1.5 years after the introduction of the flexible school start, we could control for some important confounders and address trends and complex interactions which started long before the system was changed. In addition, we also used longitudinal sleep data (duration, chronotype, social jetlag), which we had previously collected in these students^{30,31} to predict students' quarterly grades by means of linear mixed regression models.

Methods and materials

German secondary school system. The German school system varies widely between federal states, but all students start formal schooling with primary school (children aged ~6–9) followed by several possible types of mostly public secondary schools. The minimum mandatory schooling years are 9–10 years. The current study took place at a public “Gymnasium”, which is the most academic secondary school type in Germany, granting access to higher education at the university level after successful completion of the final exam (“Abiturprüfung”). Gymnasiums offer a very broad education and thus specialisations are not allowed until senior years (grades 11–13) when students can then decide to focus on specific subjects. Core subjects (e.g. maths, German and one main foreign language) cannot be deselected and—depending on the state—are examined in the final exam. About 34% of all secondary school students in Germany attend Gymnasium³².

Study site and the flexible school start system. The study took place at the Gymnasium Alsdorf (50° 53' N, 6° 10' E), a public secondary school in the West of Germany. This particular school offers daily self-study periods during which students work through a personal 5-week curriculum with a teacher and on a subject of their own choice (so-called “Dalton system”^{33,34}) to foster independent learning. This is different to most other Gymnasiums in Germany, which work with mostly teacher-centred methods and no personal curriculum.

On February 1st, 2016, the school changed permanently from a fixed start (“conventional system”) to a flexible start (“flexible system”) for 10–12th graders (senior students). In the conventional system, school started at 8 AM on most days. On a median of 1 day/week (depending on students' individual timetables), however, school started with the second period at 8:50 AM. The conventional system is a common system seen across different Gymnasiums in Germany, although the exact start time might vary slightly between schools (~7:30–8 AM). The amount of afternoon classes and the closing times vary depending on students' grade level and the school's individual schedule. Most often, students have 2–4 long school days where they attend afternoon classes finishing roughly between 1 and 5 PM.

In the flexible system, the first period (lasting 08:00–08:45 AM) was made optional for senior students to attend. Senior students could thus choose daily whether to start at 8 AM with the first *self-study period* or skip it and start at 08:50 AM instead (called “9 AM” here for convenience) with a normal teacher-centred class. On a median of 1 day/fortnight, students also had a scheduled free second period (08:50–09:50 AM), i.e. the chance to start at 10:15 AM (“>9 AM”). Given the low occurrence rate of >9 AM-starts, we did not distinguish between frequencies of 9 AM-starts and >9 AM-starts in our analyses. Skipped first periods had to be made up for within the same week, either during gap periods or after classes ended. Skipped first periods had to be made up for within the same week, either during gap periods or after classes ended.

Free first or second periods or gap periods during the day are common in secondary schools in Germany. This reflects the different choices senior students make to focus on specialised subjects which results in individualised schedules for each student. For more information on this unique flexible system, please refer to³⁰ which also includes an example timetable.

Study design. Official academic grades were obtained from the school registry for students that took part in our first wave in 2016³⁰ and the second wave in 2017³¹. While grades were provided retrospectively at the end of the schoolyear of wave 2 for the past 4 years, sleep data were collected longitudinally in both waves (Fig. 1): Wave 1 consisted of baseline sleep diary data collection (=t0) covering 3 weeks in January, 2016 (Jan 8th to 31st) in the conventional system, followed by sleep diary data collection for 6 weeks (Feb 1st to Mar 14th) in the flexible system right after its introduction on Feb 1st, 2016 (=t1). Wave 2 covered the matching photoperiod and time of

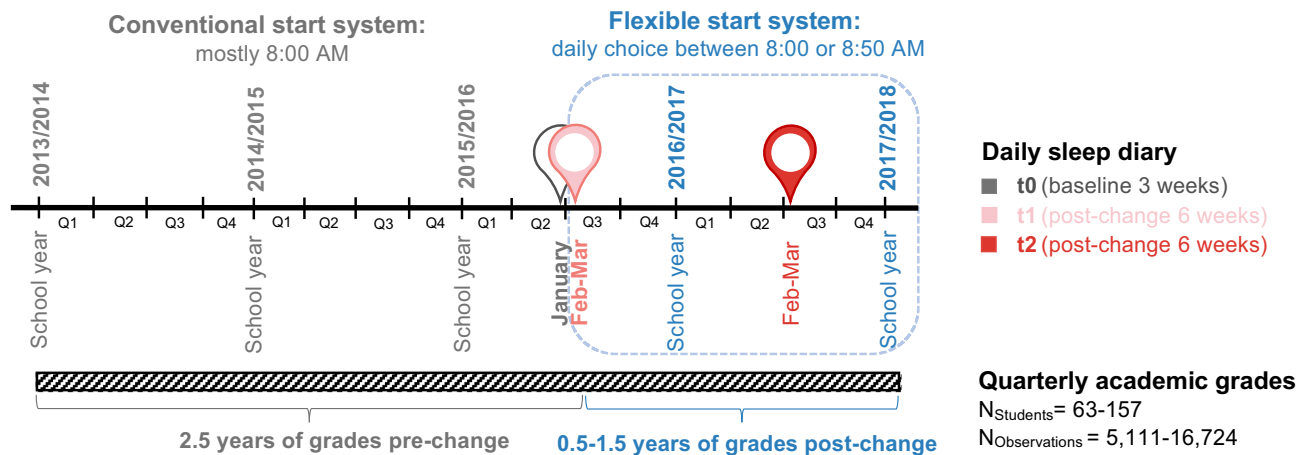


Figure 1. Study design and outcome measures. Schematic of longitudinal study design including quarterly academic grades from up to 2.5 years prior to and up to 1.5 years after the introduction of the flexible system. The same students had also provided daily sleep diary data in two waves (one baseline assessment in the conventional system and 2 time points in the flexible system as described previously³¹).

t1, lasting again 6 weeks (Feb 2nd to Mar 20th, 2017, = t2). No second baseline just before t2 was carried out since the school had remained in the flexible system. We excluded any sleep diary entries during the carnival holiday periods between February 4th–9th, 2016 and February 23rd–28th, 2017 from the analyses.

Participants. Written informed consent was obtained from all participants (or their parents/guardians if < 18y). The study was conducted according to the Declaration of Helsinki and approved by the school board, the parent-teacher association, the school's student association and the ethics committee of the Medical Faculty of the LMU Munich (#774–16). We used opportunity sampling without specific exclusion criteria. For response and attrition rates and filter criteria of sleep diary data and cohorts please refer to³¹. Academic grades were provided by the school registry from students that took part at any time during our study (t0 through t2). All included students were granted promotion to the next grade level during the study period.

Outcome measures. *Sleep diary.* We used a daily sleep diary (provided online via LimeSurvey.org) based on the μ MCTQ³⁵. Students provided sleep onset (note: not bedtime) and offset (wake time) of their past night's sleep, the type of day they woke up (school day or free day), and when they started school (8 AM, 9 AM or > 9 AM). The questionnaire did not cover any naps during the day. For detailed sleep diary descriptions, please see^{30,31}.

Academic grades. From the school registry, we obtained official quarterly grades awarded to participating students between the school year 2013/2014 through to 2016/2017. Of the 170 students from both waves qualifying for analysis (i.e. the cohorts previously described and used for sleep analyses; see³¹), 13 students had grades missing, thus resulting in a maximum sample of 157 students for the grade analyses. For the majority of these students (62%), grade data span 2.5 years in the conventional and 1.5 years in the flexible system; for those in grade level 10 at wave 2 (18%), it was 3 and 1 years, and for those at grade level 12 at wave 1 (15%) it was 2.5 and 0.5 years. The grades were provided for all academic subjects taken by a student, of which we included only 12 subjects in our analyses that most students took and assigned them to one of three disciplines: Sciences (Biology, Chemistry, Maths, Physics, Natural Sciences), Social Sciences (Geography, History), and Languages (English, German, Spanish, French, Latin). Provided grades were averages per academic quarter per academic subject over a mixture of written and oral examinations, course work and participation in class.

The school year lasted from the end of August to mid-July divided into the following quarters: quarter 1 until end of October, quarter 2 until third week of January, quarter 3 until third week of April and quarter 4 until first week of July.

In grade levels 7–10, the grading scale ranged from 1 (best) to 6 (worst) with grades ≤ 4 considered passing grades. This scale was additionally broken down into plus (+) and minus (–) for all but grade 6 (1+, 1, 1–, 2+, 2, 2– and so on). In grade levels 11 and 12, the scale ranged from 0 (worst) to 15 (best) with ≥ 4 considered passing. Both scales were combined by transforming the 1–6 scale to a 0–15 scale based on its finer plus/minus system (i.e. 15 = 1+, 14 = 1, 13 = 1–, 12 = 2+, 11 = 2, 10 = 2– and so on) and then transformed to a more universal 0–100% scale (0 = 0%, 15 = 100%, then steps of 6.66% for grades in between).

Data analysis. Analyses were performed in SPSS Statistics (IBM, versions 24 and 25) and R (versions 3.6.1 and 3.6.3) using R studio (versions 1.1.463, 1.2.1335 and 1.2.5042). Graphs were produced using the r-package *ggplot2*³⁶.

Sleep data. Daily sleep data from dairies were aggregated and taken from t2 if available (else from t1). From these aggregates, we derived the following variables as per equations below: average daily sleep duration during the week (SD_{week}), chronotype as midsleep on free days (MSF) corrected for oversleep (MSF_{sc}), social jetlag (SJL) as midsleep on free days (MSF) minus midsleep on work/school days (MSW), and frequency of ≥ 9 AM-starts. For the linear mixed Models 3a–d, we additionally calculated the absolute differences between t0 and t1 (i.e., from baseline to the flexible system during wave 1) for SD_{week} , MSF_{sc} and SJL (X change).

$$SD_{\text{week}} = (SD_{\text{ schooldays}} * 5 + SD_{\text{ free days}} * 2) / 7$$

$$MSF_{sc} = \text{SleepOnset}_{\text{free days}} + \frac{1}{2} SD_{\text{week}}$$

$$SJL = MSF - MSW$$

$$\text{Frequency of } \geq 9 \text{ AM-starts} = (n_{9 \text{ AM-starts}_{\text{flex}}} / n_{\text{ schoolday-entries}_{\text{flex}}}) * 100$$

$$X \text{ change} = x_{t1} - x_{t0}$$

Statistical analyses. Unless indicated otherwise, descriptive statistics are reported as mean \pm standard deviation and test statistics are abbreviated as follows: *t*, t-test; Pearson correlation; *rho*, Spearman rank correlation; *b*, unstandardized coefficient of linear regression or linear mixed models; $b_{\text{flex} \times \text{change}}$, unstandardized coefficient of the interaction of linear mixed models; *p*, significance level. The alpha-level was set to $p < 0.05$ for all statistical analyses. All data were tested on normality (histograms, QQ plots, Shapiro–Wilk’s test) and sphericity.

For simple grade analyses comparing accumulative grade point averages in the conventional versus the flexible system, a two-sided paired t-test was used. To this end, the grade point average was calculated as the mean grade across all subjects before or after the start-system change for each student. For more sophisticated grade analyses, we used linear mixed-effects regression models (*lme4* and *lmer test* package^{37,38} in R). In total, 4 different models (plus model variations) were calculated to answer different questions based on different fixed effects, interaction terms and subcohorts (see overview Table 1). Student ID was added as random effect to all models to incorporate unsystematic differences between individuals. In all models, the outcome (dependent variable) was quarterly grades per discipline per student; the fixed effects (independent variables) were system (conventional/flexible), gender (female/male), grade level (7–12), academic quarter (1–4), and academic discipline (Sciences/Social Sciences/Languages), all entered as categorical variables. Model 1 additionally included interaction terms between discipline and gender to assess general grade influences, Model 2 included interaction terms between school start system and gender, and system and discipline to assess system effects per discipline and gender. In Models 3, we included one of the aggregated sleep-change variables (see equation above; mean-centred) as additional fixed effects, each in interaction with system (conventional/flexible): chronotype change (Model 3a), sleep duration on schooldays change (Model 3b), social jetlag change (Model 3c) or frequency of ≥ 9 AM-starts (Model 3d). In Model 4, we instead included the absolute value of chronotype, sleep duration on schooldays, social jetlag, and frequency of ≥ 9 AM-starts for the flexible system only (from t2 if available, else from t1 to maximize sample size). Since chronotype, sleep duration on schooldays, social jetlag, and frequency of ≥ 9 AM-starts were prone to collinearity, we first assessed their correlations before adding them into the models (Supplementary Fig. S1). Only chronotype and social jetlag were at least moderately correlated ($\rho = 0.65$, $p < 0.001$; Supplementary Fig. S1), and results from models including just one of these variables each (4a–d) were essentially similar to Model 4e which included all sleep variables together (Supplementary Table S4). The variance inflation factor (*car* package in R³⁹) also indicated no problematic collinearity for Model 4e. Marginal means of model estimates were calculated using *emmeans* in R⁴⁰ for models where interactions were significant. All linear mixed models were visualised in tables using the *sjPlot* and *sjmisc* packages^{41,42} and in figures as marginal means via the *ggeffects* package⁴³ in R. Simple contrast results from interactions in linear mixed models were averaged over the levels of system or gender (depending on the model), grade level, and quarter; degrees of freedom method used was Kenward–Rogers. Pairwise comparisons were adjusted with Tukey method.

Results

In this study, we investigated whether a *flexible* school start system and concurrent changes in sleep (as previously described here^{30,31}) were associated with changes in academic grades. During a first wave of sleep assessment³⁰, the studied secondary school had changed from a conventional start system (mostly starting at 8:00 AM; baseline = t0) to a flexible start system with a *daily* choice between 8:00 and 8:50 AM (= t1; Fig. 1). The school has since maintained this system allowing for a second wave of sleep assessment after exactly one year (t2)³¹.

For the current study, we included quarterly grades of 63–157 students from these two waves irrespective of time of participation (i.e. t0/t1, t2 or all time points) (Fig. 1). The sample size varies depending on the analysis question and thus with the respective regression model calculated (see Table 1). The majority of included students were females (63–68%), were in grade levels 10 or 11 (but levels 9 and 12 were also included), and used the late-start option (“ ≥ 9 AM-use”) on about 24–28% of all recorded schooldays (see Table 2 for more cohort characteristics). In total, we analysed 5111–16,724 grades (on average 107 individual grades per person) that students received in 12 academic subjects over 2.5 years in the conventional system and 0.5 to 1.5 years in the flexible system (Fig. 1). Grades were provided by the school registry and transformed to a 0–100% scale and

	Model 1	Model 2	Model 3a–d	Model 4a–c
Outcome	Official grades (per quarter and academic subject)	Official grades (per quarter and academic subject)	Official grades (per quarter and academic subject)	Official grades (per quarter and academic subject)
Aim	General effects	System effects	Effects of sleep changes	Effects of sleep in flexible system only
Fixed effects	System (conv/flex)	System (conv/flex)	System (conv/flex)	–
	Gender	Gender	Gender	Gender
	Grade level	Grade level	Grade level	Grade level
	Academic quarter	Academic quarter	Academic quarter	Academic quarter
	Academic discipline	Academic discipline	Academic discipline	Academic discipline
			Change ^a in	–
			a. Chronotype	a. Chronotype ^c
			b. Sleep duration	b. Sleep duration ^{c,d}
Interactions	Gender* Academic discipline	System* Academic discipline System* Gender	System* Chronotype change/Sleep duration change/ Social jetlag change / ≥ 9 AM-use	–
Random intercept	ID	ID	ID	ID
Sample	Waves 1 & 2	Waves 1 & 2	Wave 1 only	Waves 1 & 2
N	157	157	63	129
Number of observations	16,724	16,724	6683	5111
Data span	4 years: 2.5 y conv & 1.5 y flex	4 years: 2.5 y conv & 1.5 y flex	4 years: 2.5 y conv & 1.5 y flex	1.5 years: Only flex

Table 1. Overview of linear mixed model analyses on official, quarterly grades. Four different models (and several model variations) were calculated, each with a different aim and including appropriate predictors (fixed effects) and interaction terms. All models included ID as a random intercept to incorporate random inter-individual differences. *conv* conventional school start system, *flex* flexible school start system. ^aChange refers to the absolute difference between the respective variable at t1 minus t0 (baseline). Positive values indicate higher numbers at t1. ^bSince the exact frequency of 9 AM-starts during baseline (t0) is not known, 9 AM-use was added as an absolute value rather than the change from t0 to t1. Students attended school at ≥ 9 AM at a median of 1 day per week in the conventional system. ^cFrom t2 if possible, else from t1. ^dDuration on schoolday.

labelled (not aggregated) as Languages, Sciences, or Social Sciences. Median grades were 53–60% in Languages, 60% in Sciences, and 60–67% in Social Sciences (Table 2). To complement the analyses, we also used several of the existing sleep variables (chronotype expressed as MSF_{sc} , sleep duration, social jetlag) in the flexible system and their respective change (delta) from baseline (t0) to the flexible system (t1) as well as the frequency of ≥ 9 AM-use in several model specifications (Tables 1 and 2).

School start system showed no systematic effect on academic grades overall. First, we investigated whether the flexible system allowed students to increase their grades without considering sleep variables. At first sight, a simple comparison of overall grades yielded a small but statistically significant improvement in grade point average from 58.2% (± 2.1 SD) in the conventional to 59.6% (± 2.0 SD) in the flexible system (Fig. 2a; $t[154] = -2.15$, $p = 0.033$, $d_z = 0.173$). However, attributing this improvement to the flexible system is likely unwarranted. As outlined in the introduction, grades are influenced by a multitude of factors, thus comparisons that do not account for these can be misleading. We therefore applied linear mixed-effects regression models to adjust for potential confounders and including random intercept for ID to account for inter-individual differences (Table 1). When incorporating gender, grade level (i.e. indirectly age), academic quarter, and discipline in addition to school start system in the analysis, the flexible system showed no systematic relationship with students' grades (Fig. 2b; $b = -0.10$, $p = 0.815$, Model 1, Supplementary Table S1), hence the flexible system was not associated with students receiving better or worse grades overall in our sample ($n_{\text{students}} = 157$).

Grades varied systematically with grade level, academic quarter, discipline and gender. But what drives better grades in the unadjusted comparison if not the flexible system itself? The same factors that we adjusted for in the regressions also stood out as major predictors (Model 1, Supplementary Table S1, $n_{\text{students}} = 157$): Students in 12th grade (the last school year) did consistently better compared to their peers across all other grade levels—a sort of “leavers effect” that has already been observed before (Fig. 2c, $b = 3.44$, $p < 0.001$)⁴⁴. Moreover, we found that students enjoyed a bump in grades in the last quarter of the school year with an estimated improvement of 2.3 percentage points compared to the first quarter (Fig. 2d; $b = 2.34$, $p < 0.001$). The combination of these two effects might explain the statistically significant improvement observed in the unadjusted comparison:

	Sample incl. in Models 1 and 2	Sample incl. in Model 3	Sample incl. in Model 4
Participants			
Total			
n	157	63	129
Females			
n (%)	109 (69%)	40 (63%)	88 (68%)
Grade level			
n (%) per level 9th/10th/11th/12th	29/50/52/26 (18/32/33/17%)	0/25/21/17 (0/40/33/27%)	24/43/45/17 (19/33/35/13%)
Academic grades per discipline on a scale from 0% (worst)–100% (best)			
Languages			
Median (IQR)	53% (40–73)	53% (47–70)	60% (46–73)
Sciences			
Median (IQR)	60% (47–73)	60% (47–73)	60% (46–73)
Social Sciences			
Median (IQR)	60% (53–73)	60% (53–73)	67% (53–73)
		Change (Δ t1-t0)	t2 if available, else t1
Sleep variables			
Chronotype (MSF _{sc} : h or time)			
Mean (SD, range)	Not used in model	0.2 (0.7, – 1.6–2.2)	4:40 (1.0, 2:10–8:35)
Social jetlag (h)			
Mean (SD, range)	Not used in model	– 0.3 (0.7, – 2.0–1.9)	2.0 (0.8, 0.5–6.0)
Sleep duration (h)			
Mean (SD, range)	Not used in model	0.1 (0.5, – 1.3–1.1)	7.2 (0.8, 5.2–9.0)
Frequency of attending school later			
≥ 9 AM-use			
Median (IQR)	Not used in model	24% (14–46)	28% (10–52)

Table 2. Composition of study cohorts. Displayed are cohort characteristics after standard filter criteria. Grades were provided by the school registry. Grade levels are taken from t1. *n* number of individuals, *SD* standard deviation, *IQR* interquartile range, *conv.* conventional.

the flexible system replaced the conventional system mid-year between quarter 2 and 3, so quarter 4 and higher grade levels were overrepresented in the flexible system, which the t-test could not account for.

The mixed models also revealed other strong systematic influences on grades in our sample. Firstly, we observed a clear difference between the disciplines: students performed generally best in Social Sciences, followed by Sciences and then Languages (Model 1, Supplementary Table S1). Post-hoc tests (Fig. 2e) showed that these differences were highly significant for both genders (all $p < 0.001$; post-hoc to Model 1, Supplementary Table S2), except for girls' grades in Sciences and Social Sciences, which were indistinguishable ($b = -0.47$, $p = 0.3895$; post-hoc to Model 1, Supplementary Table S2).

Female gender has been reported as another driving force for higher grades⁴⁵. However, girls in our sample did not outperform boys overall (Model 2, Supplementary Tables S1 and S2). Girls were significantly better in Languages (Fig. 2e; $b = 4.72$, $p = 0.0284$; post-hoc to Model 1, Supplementary Table S2), while boys surpassed them in the Social Sciences ($b = -3.31$, $p = 0.1269$; post-hoc to Model 1, Supplementary Table S2), and both genders did equally well in Sciences ($b = -0.00$, $p = 0.9915$; post-hoc to Model 1, Supplementary Table S2).

The flexible system was linked with subtle improvements in languages and subtle drops in social sciences grades.

Although we did not find evidence that the flexible system was linked with better grades overall (Model 1, see above), the flexible system might be linked with grade improvements in certain disciplines and genders. To assess this, we looked at the interaction between (1) school start system and discipline, as well as (2) school start system and gender in a second model (Model 2; Supplementary Table S1, $n_{\text{students}} = 157$). Neither females nor males significantly improved their overall grades from the conventional to the flexible system (Fig. 2f; post-hoc to Model 2, Supplementary Table S2). In terms of discipline effects, we found that grades in Social Sciences slightly dropped ($b = 1.26$, $p = 0.0384$; post-hoc to Model 2, Supplementary Table S2), Science grades remained unchanged ($b = -0.07$, $p = 0.8849$; post-hoc to Model 2, Supplementary Table S2), and Language grades slightly improved ($b = -1.30$, $p = 0.0168$, post-hoc to Model 2, Supplementary Table S2) in the flexible system. Notably, these changes were subtle but reduced the grade differences between the academic disciplines (Fig. 2g, Supplementary Table S2). These small changes in opposite directions likely explain the absence of a net effect of the flexible system on overall grades.

Improvements in chronotype, sleep duration, and social jetlag did not systematically improve grades.

What was the role of sleep parameters on grade developments? We speculated that students who showed greater improvements in the flexible system (i.e., advanced chronotype, lengthened sleep duration, and

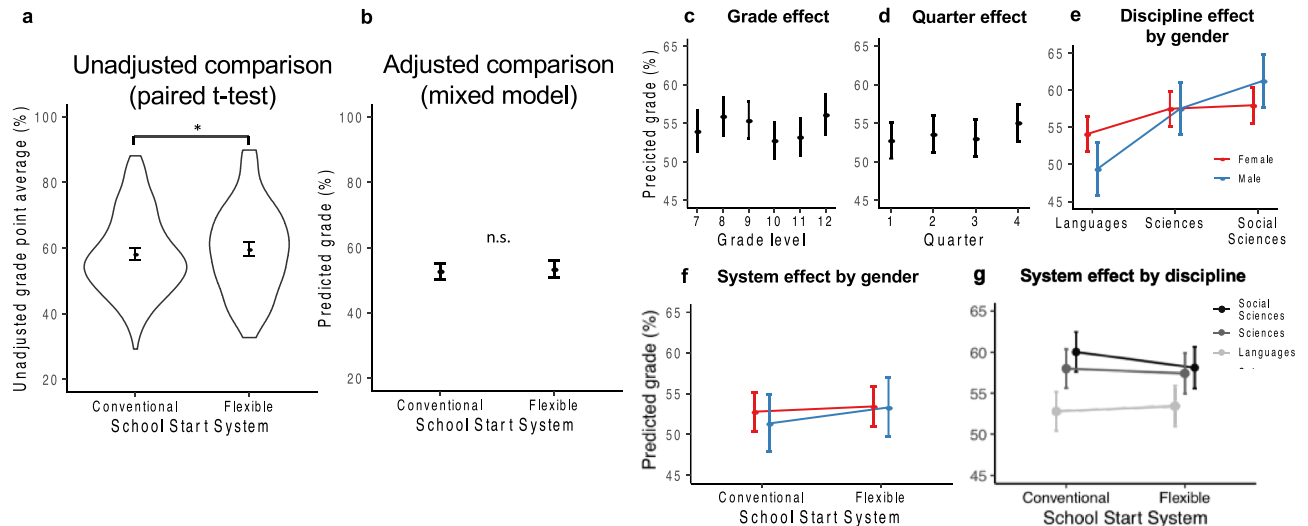


Figure 2. Longitudinal analysis of official quarterly grades—effects of school start system and general predictors. Quarterly grades (0–100%) from 12 academic subjects of 3 disciplines for 4 years i.e., for most students this was 2.5 years before and 1.5 years after the flexible school start was introduced ($n = 157$ students; 16,724 grades; 107 grades per student on average). **(a)** Simple, unadjusted comparison of average grades across all disciplines in the conventional and the flexible school start system via paired t-test ($n_{ID} = 157$). Shown are mean and 95% CI within the raw data distribution (violin plots). The apparent grade improvement in the flexible system was not confirmed in linear mixed models. **(b–g)** Visualization of mixed-model-determined influences on grades. Plots show marginal means from Models 1 and 2 (Supplementary Table S1), i.e. the estimated grade and 95% CI for the reference category (female student, class level 10, quarter 1, languages, conventional system). Statistical significance is indicated in **(b)**, results of more complex cases can be found in the text and Supplementary Tables S1 and S2. **(b)** Effect of school start system (Model 1). **(c)** Effect of grade level (Model 1). **(d)** Effect of academic quarter (Model 1). **(e)** Effect of academic discipline by gender (Model 1). **(f)** Effect of school start system by gender (Model 2). **(g)** Effect of school start system by academic discipline (Model 2).

lowered social jetlag) also received better grades in the flexible system. Thus, we computed changes in sleep from t_0 (baseline) to t_1 for each student in the subpopulation of students with sleep parameters during these time points ($n = 63$; Table 2). On average, these students showed a small delay in chronotype of 0.2 h (MSF_{sc} , range = -1.6 – 2.2 h), a small reduction in social jetlag of 0.3 h (-2.0 – 1.9 h) and a small sleep gain of 0.1 h (-1.3 – 1.1 h) but large inter-individual differences as reflected by the ranges. Adding these parameters separately into a third model (Models 3a–c, Supplementary Table S3, Fig. 3a), we found that neither changes in chronotype (flex*chronotype change: $b = 0.10$, $p = 0.845$, Fig. 3a,b) nor changes in sleep duration (flex*sleep duration change: $b = -0.77$, $p = 0.352$, Fig. 3a,c) were systematically associated with changes in grades. Surprisingly, however, students who increased their social jetlag in the flexible system obtained slightly better grades in the flexible system (flex*social jetlag change: $b = 1.28$, $p = 0.027$; Fig. 3a,d), which was contrary to our hypothesis. Therefore, our analyses in this subsample suggest that sleep improvements experienced immediately after transitioning to the flexible system did not result in detectable higher academic achievement.

If not linked to sleep improvements, were grades nonetheless linked with the choice of more later school starts? The results of Model 3d calculated to answer this question suggests that higher 9 AM-use was associated with worse grades in the conventional system ($b = -3.04$, $p = 0.015$), a link reversed partly—albeit not significantly—in the flexible system (flex*9 AM-use: $b = 0.59$, $p = 0.101$; Fig. 3a,e and Supplementary Table S3). Hence, students who made high use of the late-start option in the flexible system (group median was 24% of schooldays or 1 day/week; Table 2) were predominantly students obtaining lower grades, but they tended to benefit at least slightly from more later starts.

No systematic effects of chronotype, social jetlag and sleep duration on grades. Lastly, we investigated if we could find absolute effects of sleep variables (chronotype, social jetlag, sleep duration) and ≥ 9 AM-use on grades in the flexible system (Model 4, $n = 129$ students). Table 2 shows descriptive statistics of these three variables in this cohort (notice the wide range): chronotype or midsleep time on weekends corrected for oversleep was on average at 4:40 AM (range = 2:10–8:35), social jetlag was on average 2.0 h (0.5–6.0 h) and students slept on average 7.2 h (5.2–9.0 h). In contrast to what we had expected, none of the sleep parameters showed any significant link with grades, independent of whether they were added separately into the model (Model 4a–d; Supplementary Table S4) or together (Model 4e; Fig. 3f, Supplementary Table S4). Our results thus indicate that late chronotypes in our sample were not worse off compared to their early peers (Fig. 3f,g) and that longer sleep duration in the flexible system did not improve grades received in the flexible system (Fig. 3f,h). Similarly, social jetlag did not hamper grades to such a degree that we could detect an effect (Fig. 3f,i). Furthermore, frequency of attending school later was at a median of 1.5 days per schoolweek in this sample (28% of

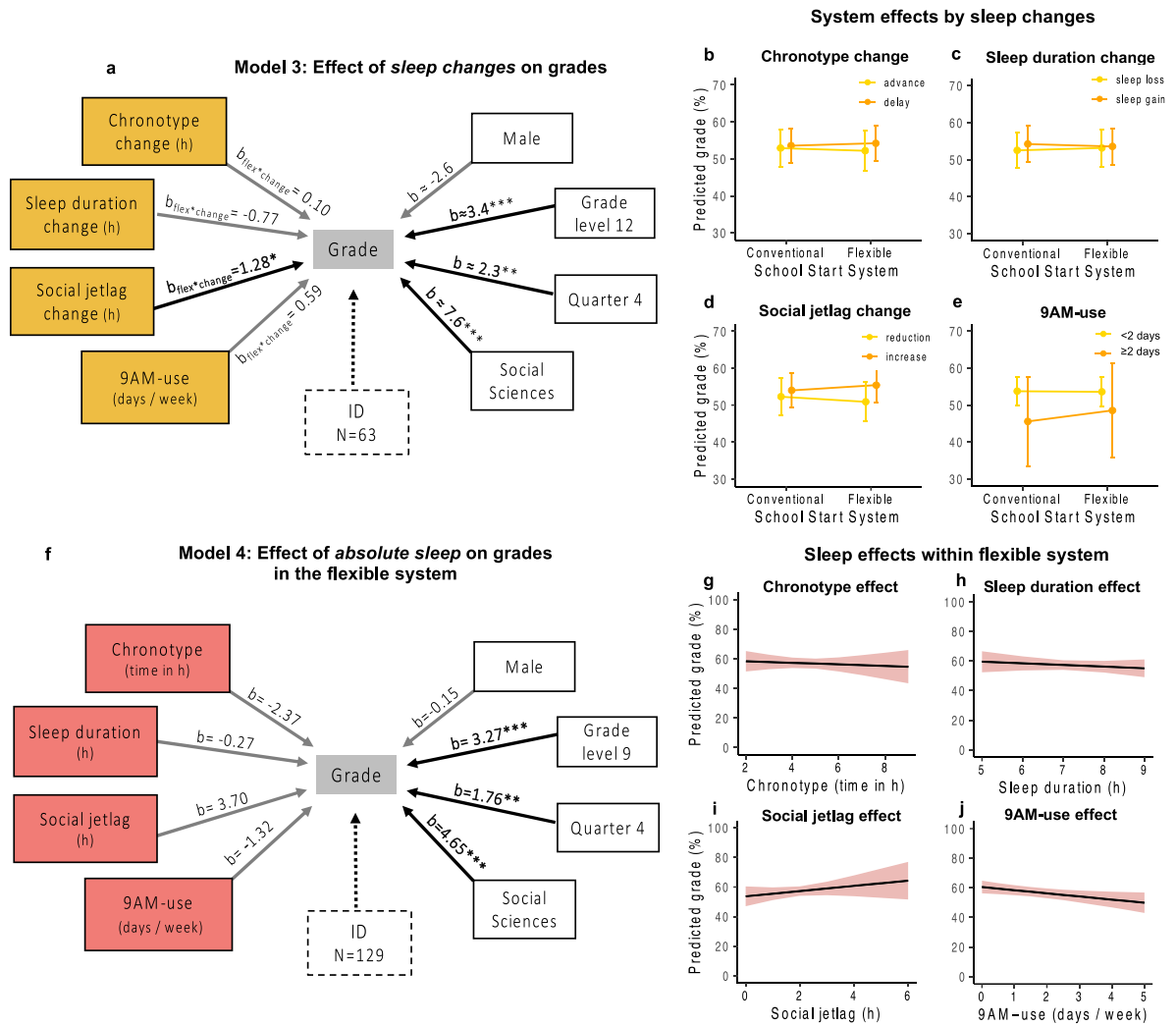


Figure 3. Longitudinal analysis of official quarterly grades—effects of sleep and 9 AM-use. Results from linear mixed model analyses of quarterly grades (0–100%) considering sleep variables and the frequency of ≥ 9 AM-starts (see Table 2 for sample descriptions). **(a,f)** Schematic of the structure and results from Models 3 and 4 (Supplementary Tables S3 and S4) showing the outcome, official quarterly grades (center), all predictors (black-framed boxes), the statistical significance of their effect (arrows; black: $p < 0.05$, grey: $p \geq 0.05$), the unstandardized regression coefficients (b-values) and ID as random intercept (dashed box). General predictors (white) are categorical variables, so the levels with the highest impact are shown compared to their reference (female, grade level 10, quarter 1, languages). b-values are approximate in **(a)**, indicated by \approx , as representing results from Models 3a–d. **(a)** Effect of changes in sleep and of 9 AM-use on grade improvements from the conventional to the flexible system. Summarized results from Models 3a–d ($n_{ID} = 63$; Supplementary Table S3) where each model variation included a different yellow predictor in interaction with school start system (conventional/flexible; $b_{flex*change}$) to model effects of sleep changes on grade changes. **(b–e)** Visualization of the yellow interaction effects from **(a)** via marginal means, i.e. grade estimates and 95% CI for the reference (female student, class level 10, quarter 1, languages) and categorical splits in the continuous sleep change variables to facilitate display. The effect of school start system on grades by **(b)** chronotype change (advance/delay), **(c)** sleep duration change (sleep loss/sleep gain), **(d)** social jetlag change (reduction/increase) from the conventional to the flexible system, and by **(e)** the frequency of 9 AM-use (<2 days/ ≥ 2 days) in the flexible system. **(f)** Effect of absolute sleep characteristics on grades in the flexible system. Summarized results from Model 4e ($n_{ID} = 129$; Supplementary Table S4) predicting grades only for the flexible system, i.e., 1.5 years post-change, including the red sleep predictors in one common model after running separate Models (4a–d) to check for collinearity. **(g–j)** Visualization of the red effects from **(f)** via marginal means, i.e. grade estimates and 95% CI for the reference (female student, class level 10, quarter 1, languages). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

schooldays, IQR = 10–52%; Table 2), and we found that per every additional day a student chose to go to school later, grade estimates non-significantly decreased by 1.32 in the full Model 4e ($p = 0.272$, Fig. 3f,j; but see the single Model 4d in the Supplementary Table S4: $b = -2.12$, $p = 0.022$). Although the interpretation is slightly differ-

ent, this result tallies with the above finding from Model 3: At first sight, it looks as if attending school later more often would prevent students from getting better grades, but we argue that most likely it is the other way around; students who receive worse grades also liked to attend school later more often when they had the chance to do so. Overall, we could not show that chronotype and social jetlag negatively influenced grades, and it seemed as if mainly students who previously achieved lower grades in our sample liked to use the ≥ 9 AM option.

Discussion

Adolescence is a decisive time in life for teenagers around the world. Teenagers undergo many cognitive, emotional and brain structural changes that also shape their risk-taking behaviour, learning capacities and motivation to attend school^{46,47}. A prominent change also occurs in their daily sleep–wake behaviour: teenagers tend to phase-delay their sleep–wake behaviour, which essentially means that they become night-owls^{5,48–52}. This delayed phase, however, clashes with early school starts seen across many countries, thus cutting sleep short in the morning hours during the school week. Apart from many other negative (health) consequences^{10,11,53–56}, short and low-quality sleep as well as sleepiness likely influences academic success^{20,57}, which in turn is an important determinant of future career trajectories⁵⁸. Since sleep restrictions and poor sleep habits are more severe in minority groups and disadvantaged students^{59,60}, addressing this problem is key to closing the achievement gap between social groups. However, the evidence is not conclusive whether delayed school start times can ameliorate this pressing health and performance problem. Many previous studies suffer from study design limitations, outcome variables are not comparable, and long-term studies that track individuals over time are rare²⁸. Here, we studied whether a novel timetabling system—a daily chosen flexible school start—has the potential to improve academic grades via improved sleep.

In our study, we found that the flexible system was only associated with higher grades at first sight. When not adjusting for confounding factors, we observed a small improvement of grades in the flexible system, which would be in line with some previous studies e.g.,^{55,56}. However, we argue that such simple pre-post analysis of aggregated grades is not suited to answer this complex question—although this has frequently been done using cross-sectional data. Studies on grades that performed proficient analyses, such as mixed regression models⁴⁴, quantile regression models⁶¹ or difference-in-difference approaches^{62–64} accounting for available confounders provided mixed results and mostly small effect sizes (for a systematic review see²⁸). Nonetheless, positive effects of delayed school start times on academic achievement have been widely proclaimed—bound to raise falsely high expectations in parents and teachers. When we considered grade level, discipline and quarter in mixed model analyses, we found that the flexible system was clearly not associated with overall grade improvements except for subtle increases in Languages and subtle decreases in Social Sciences. In fact, the “confounders” weighed much stronger in our sample than any school start system effect on individual disciplines: graduating students did constantly better, highest grades were given in the final quarter of the year, and students were most successful in Social Sciences. Furthermore, the interplay between gender, discipline and school start system on grades is complex.

Importantly, we also did not find any expected relationships between chronotype, social jetlag, or sleep duration with grades in our sample. Neither changes in these sleep parameters from the conventional to the flexible system nor their absolute values in the flexible system showed any link with grades—except for changes in social jetlag. Surprisingly, an increase in social jetlag, not a decrease, in the flexible system was predictive of higher grades in the flexible system. We have not been able to identify obvious explanations for this finding in exploratory analyses, except for the fact that weekend sleep was much more variant and backed by fewer data points than schoolday sleep, pointing towards a potential chance finding. A likely explanation for our null-finding for the other sleep parameters is a possible lack of power in our sample of 157 students (even though we have > 16,000 longitudinal grades) given the small effect sizes previously identified (ranging around < 0.1 SD; see²⁸). A second possibility is that the time frame students were exposed to the new system was too short (exposure length) or that the delay was too little or infrequent (dose) in our study. Furthermore, sleep variables obtained at discrete study points might not be reflective of sleep during the other academic quarters or years. Thus, we cannot preclude that we missed a subtle effect in our sample but any such effect is likely extremely small. This is also in line with several other studies that were unable to find any effect or meaningful improvements²⁸.

Importantly, the fact that we did not detect systematic improvements in students' grades does not mean that there were no improvements in learning. There is a substantial body of evidence supporting that both acute and chronic sleep loss compromises alertness, cognitive performance and memory, and reduces engagement to perform well (performance effort)^{15,65,66}. Thus, improving sleep in sleep-deprived teenagers is very likely to improve their learning^{67–69}. In addition, one could speculate that the flexibility and the thus putatively increased self-responsibility and self-determination of students in the flexible system, paired with the reported increase in motivation on later days, may also further improve learning. The question is whether better learning mediated by improved sleep also translates into better grades—and how much sleep improvement is needed and within what timeframe.

Additionally, students' learning is strongly affected by many factors beyond those captured in our study or those of others on this topic. Models of teaching and learning include several nested factors, such as the individual student (e.g. motivation and prior knowledge), the individual teacher (professional competence^{70,71}), the learning environment (e.g. socio-economic status or native language) or factors of instruction (generic and subject-specific instructional quality) but also class-level factors (learning atmosphere or class mates)⁷². Of these, especially instruction and teacher-level factors greatly influence students' learning^{73,74}. Furthermore, grades are inherently suboptimal measures of students' academic performance, as teachers also include other factors such as compliance, effort, attitude, or behaviour in their assessment⁷⁵.

Therefore, it may be a big ask and possibly naive to expect grades to improve noticeably and within a few months after delays or a flexible system have been affected. Rather, we should acknowledge students' maintained achievements under potentially less effort and improved learning capacities (this needs to be assessed in future studies) in addition to the gift of more sleep and better well-being. Indeed, teachers at the studied school reported perceiving students as more alert and motivated and tardiness rates as decreased (personal communication). On the other hand, grades still determine future career trajectories and open doors to higher education in many countries⁷⁶. In this sense, they do have a greater importance for careers—at least early on—than other measures of performance, such as standardised or non-standardised tests in class, which might be more valid for measuring academic performance under certain conditions. Additionally, despite all described influences, grades—as indicators of prior knowledge—seem to be the best predictor for achievement in university courses⁷⁷.

Our study has several limitations that have not yet been mentioned. The analyses are based on quarterly aggregates of grades, so we could not consider the type of grades (oral, written, etc.) and their weights thus prohibiting detailed analyses on grade types and their possible competence aspects (e.g. knowledge, attitude, ability).

Additionally, we did not obtain information about teachers' competence, their instructional quality or classroom atmosphere but accounted for gender, quarter, grade level, and discipline—factors that are often overlooked in the field. We also lacked socio-demographic information, which likely influence grades, such as the socioeconomic status (SES), parents' education or ethnicity of students and their parents⁷⁸. The vast majority of students included in this study were Caucasian by observation, so we had low variation with regards to ethnicity. Furthermore, the school studied was a Gymnasium, which is the most academic secondary school type in Germany leading directly to higher education. Therefore the generalisability of our results to non-Gymnasium secondary students in Germany (~65%³²) is limited. Lastly, we did not collect objective measures of cognitive performance through cognitive test batteries but asked students to self-evaluate their quality of study and concentration levels³¹. We also did not collect information on napping behaviour, which might have led to an underestimation of sleep duration overall thus underestimating the positive effects of flexible start times on sleep.

In conclusion, we highlight that current early school start times around the globe are detrimental for sleep and health and likely do not allow students to excel as much as they could. Many studies have shown positive effects on sleep or well-being, when school starts were delayed^{21–24} or in systems where school already starts much later, such as in Uruguay or Argentina^{79,80}. Thus, it seems fair to argue that later starts are beneficial for students in terms of health and well-being. These factors form a profound basis for good academic achievement but there are also numerous other factors that play into this and possibly mask positive effects: for example, teachers might not perform at their best later during the day or adjust their grading under the new bell times to achieve normal distributions of performance; the dose of delay might need to be higher (i.e. more delay or more uptake of later starts in a flexible system) and exposure time might need to be longer until an effect emerges, or grades are insensitive to this kind of intervention, to name only a few. But despite these complications, it should be emphasized that students can maintain their grades *in addition to* better sleep and well-being—a central and very important achievement in its own right. Indeed, students might need to spend less time on their homework, learn more easily or show improved sustained attention while still achieving similar but not improved grades. This question should be addressed in future research, e.g. by distributing questionnaires on time use data in general but also time spent on homework, learning and other academic activities, or running cognitive and attention tests in the field (such as the psychomotor vigilance task assessed for example with the NASA PVT + app⁸¹). Furthermore, more studies are required on how to harness the unique advantages of flexible start systems, such as promoting students' responsibility, choice and investment, for optimal sleep and learning gains.

Data availability

Open access sharing of data is not possible due to consent forms that prohibit online deposition of data. We implemented this to secure students' privacy since students and teachers were well acquainted with each other and might have identified specific participants. Data are available from the corresponding authors upon reasonable request.

Code availability

We did not develop any custom code or algorithms for data analyses. The code for grade analyses can be found here on github (<https://github.com/annambiller/School-study/blob/13bfb38c35332fe285b3cab738b621a4083df7f6/Linear%20mixed%20model%20code%20for%20grade%20analyses.txt>).

Received: 16 September 2021; Accepted: 25 January 2022

Published online: 24 February 2022

References

1. Kopasz, M. *et al.* Sleep and memory in healthy children and adolescents: A critical review. *Sleep Med. Rev.* **14**, 167–177 (2010).
2. Jenni, O. G. & Carskadon, M. A. Spectral analysis of the sleep electroencephalogram during adolescence. *Sleep* **27**, 774–783 (2004).
3. Huber, R. High-density sleep EEG recordings during adolescence. *J. Sleep Res.* **17**, 1–39 (2008).
4. Giedd, J. N. Structural magnetic resonance imaging of the adolescent brain. *Ann. N. Y. Acad. Sci.* **1021**, 77–85 (2004).
5. Roenneberg, T. *et al.* A marker for the end of adolescence. *Curr. Biol.* **14**, 1038–1039 (2004).
6. Carskadon, M. A., Acebo, C. & Jenni, O. G. Regulation of adolescent sleep: Implications for behavior. *Ann. N. Y. Acad. Sci.* **1021**, 276–291 (2004).
7. Jenni, O. G., Achermann, P. & Carskadon, M. A. Homeostatic sleep regulation in adolescents. *Sleep* **28**, 1446–1454 (2005).
8. Wittmann, M., Dinich, J., Meroow, M. & Roenneberg, T. Social jetlag: Misalignment of biological and social time. *Chronobiol. Int.* **23**, 497–509 (2006).
9. Roenneberg, T., Pilz, L. K., Zerbini, G. & Winnebeck, E. C. Chronotype and social jetlag: A (self-) critical review. *Biology (Basel)* **8**, 1–19 (2019).

10. Garaulet, M. *et al.* Short sleep duration is associated with increased obesity markers in European adolescents: Effect of physical activity and dietary habits. The HELENA study. *Int. J. Obes.* **35**, 1308–1317 (2011).
11. Mullington, J. M., Haack, M., Toth, M., Serrador, J. M. & Meier-Ewert, H. K. Cardiovascular, inflammatory, and metabolic consequences of sleep deprivation. *Prog. Cardiovasc. Dis.* **51**, 294–302 (2009).
12. Fitzgerald, C. T., Messias, E. & Buysse, D. J. Teen sleep and suicidality: Results from the youth risk behavior surveys of 2007 and 2009. *J. Clin. Sleep Med.* **7**, 351 (2011).
13. Raniti, M. B. *et al.* Sleep duration and sleep quality: Associations with depressive symptoms across adolescence. *Behav. Sleep Med.* <https://doi.org/10.1080/15402002.2015.1120198> (2017).
14. Baum, K. T. *et al.* Sleep restriction worsens mood and emotion regulation in adolescents. *J. Child Psychol. Psychiatry Allied Discip.* **55**, 180–190 (2014).
15. Killgore, W. D. S. *et al.* Sleep deprivation reduces perceived emotional intelligence and constructive thinking skills. *Sleep Med.* **9**, 517–526 (2008).
16. Horne, J. A. Sleep loss and 'divergent' thinking ability. *Sleep* **11**, 528–536 (1988).
17. Randazzo, A. C., Muehlbach, M. J., Schweitzer, P. K. & Walsh, J. K. Cognitive function following acute sleep restriction in children ages 10–14. *Sleep* **21**, 861–868 (1998).
18. Harrison, Y. & Horne, J. A. Sleep deprivation affects speech. *Sleep* **20**, 871–877 (1997).
19. Harrison, Y. & Horne, J. A. Sleep loss impairs short and novel language tasks having a prefrontal focus. *J. Sleep Res.* **7**, 95–100 (1998).
20. Dewald, J. F., Meijer, A. M., Oort, F. J., Kerkhof, G. A. & Bögels, S. M. The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep Med. Rev.* **14**, 179–189 (2010).
21. Bowers, J. M. & Moyer, A. Effects of school start time on students' sleep duration, daytime sleepiness, and attendance: A meta-analysis. *Sleep Heal.* **3**, 423–431 (2017).
22. Minges, K. E. & Redeker, N. S. Delayed school start times and adolescent sleep: A systematic review of the experimental evidence. *Sleep Med. Rev.* **28**, 82–91 (2016).
23. Marx, R. *et al.* Later school start times for supporting the education, health, and well-being of high school students. *Campbell Syst. Rev.* **13**, 1–99 (2017).
24. Widome, R. *et al.* Association of delaying school start time with sleep duration, timing, and quality among adolescents. *JAMA Pediatr.* **174**, 697–704 (2020).
25. Walker, M. P. & Stickgold, R. Sleep, memory, and plasticity. *Annu. Rev. Psychol.* **57**, 139–166 (2006).
26. Diekelmann, S. & Born, J. The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010).
27. Tarokh, L., Short, M., Crowley, S. J., Fontanellaz-Castiglione, C. E. G. & Carskadon, M. A. Sleep and circadian rhythms in adolescence. *Curr. Sleep Med. Reports* **5**, 181–192 (2019).
28. Biller, A. M., Meissner, K., Winnebeck, E. C. & Zerbini, G. School start times and academic achievement: A systematic review on grades and test scores. *Sleep Med. Rev.* <https://doi.org/10.1016/j.smrv.2021.101582> (2021).
29. Mazzilli Louzada, F. & Isabel Ribeiro Pereira, S. Adolescents' sleep/wake patterns and school schedules: Towards flexibility. *Biol. Rhythm Res.* **50**, 78–84 (2019).
30. Winnebeck, E. C. *et al.* Later school start times in a flexible system improve teenage sleep. *Sleep* **43**, 1–17 (2020).
31. Biller, A. M., Molenda, C., Zerbini, G., Roenneberg, T. & Winnebeck, E. C. One year later: Longitudinal effects of flexible school start times on teenage sleep and subjective psychological outcomes. *Scientific Reports* (2021). <https://doi.org/10.1038/s41598-022-06209-4>.
32. Statistisches Bundesamt. Schulen auf einen Blick. 1–56 https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Publikationen/Downloads-Schulen/broschuere-schulen-blick-0110018169004.pdf?__blob=publicationFile (2016).
33. Parkhurst, H. *Education on the Dalton Plan* (E.P. Dutton & Company, 1922).
34. Dalton International. <https://daltoninternational.org/>
35. Ghotbi, N. *et al.* The μ MCTQ: An ultra-short version of the munich chronotype questionnaire. *J. Biol. Rhythms* **35**, 98–110 (2020).
36. Wickham, H. *et al.* R package ggplot2: Elegant Graphics for Data Analysis. Version 3.2.1. <https://cran.r-project.org/package=ggplot2> (2019).
37. Bates, D., Mächler, M., Bolker, B. & Walker, S. R package: Fitting Linear Mixed-Effects Models Using lme4. Version 1.1-8. (2015).
38. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. R package lmerTest: Test in Linear Mixed Effects Models. Version 3.0-1. (2018).
39. Fox, J. *et al.* R package car: Companion to Applied Regression. Version 3.0-8. (2020).
40. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. R package emmeans: Estimated Marginal Means, aka Least-Square Means. Version 1.4.6. (2020).
41. Lüdtke, D., Bartel, A., Schwemmer, C., Powell, C. & Djalovski, A. R package sjPlot: Data Visualization for Statistics in Social Science. Version 2.8.4. (2020).
42. Lüdtke, D., Giné-Vásquez, I. & Bartel, A. R package sjmisc: Data and Variable Transformation Functions. Version 2.8.4. (2020).
43. Lüdtke, D. & Aust, F. R package ggeffects: Create Tidy Data Frames of Marginal Effects for 'ggplot' from Model Outputs. Version 0.15.0. (2020).
44. Thacher, P. V. & Onyper, S. V. Longitudinal outcomes of start time delay on sleep, behavior, and achievement in high school. *Sleep* **39**, 271–281 (2016).
45. Voyer, D. & Voyer, S. D. Gender differences in scholastic achievement: A meta-analysis. *Psychol. Bull.* **140**, 1174–1204 (2014).
46. Crone, E. A. & Dahl, R. E. Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nat. Rev. Neurosci.* **13**, 636–650 (2012).
47. Dahl, R. E. & Lewin, D. S. Pathways to adolescent health: Sleep regulation and behavior. *J. Adolesc. Heal.* [https://doi.org/10.1016/S1054-139X\(02\)00506-2](https://doi.org/10.1016/S1054-139X(02)00506-2) (2002).
48. Crowley, S. J., Wolfson, A. R., Tarokh, L. & Carskadon, M. A. An update on adolescent sleep: New evidence informing the perfect storm model. *J. Adolesc.* <https://doi.org/10.1016/j.adolescence.2018.06.001> (2018).
49. Crowley, S. J., Cain, S. W., Burns, A. C., Acebo, C. & Carskadon, M. A. Increased sensitivity of the circadian system to light in early/mid-puberty. *J. Clin. Endocrinol. Metab.* **100**, 4067–4073 (2015).
50. Carskadon, M. A. Evolution of sleep and daytime sleepiness in adolescents. *Sleep/wake Disord. Nat. Hist. Epidemiol. Long-Term Evol.* (1983).
51. Hagenauer, M. H., Perryman, J. I., Lee, T. M. & Carskadon, M. A. Adolescent changes in the homeostatic and circadian regulation of sleep. *Dev. Neurosci.* **31**, 276–284 (2009).
52. Carskadon, M. A., Wolfson, A. R., Acebo, C., Tzischinsky, O. & Seifer, R. Adolescent sleep patterns, circadian timing, and sleepiness at a transition to early school days. *Sleep* **21**, 871–881 (1998).
53. Åkerstedt, T. *et al.* Sleep duration and mortality: Does weekend sleep matter?. *J. Sleep Res.* **28**, 1–11 (2019).
54. Bonnet, M. H. & Arand, D. L. We are chronically sleep deprived. *Sleep* **18**, 908–911 (1995).
55. Beebe, D. W. Cognitive, behavioral, and functional consequences of inadequate sleep in children and adolescents. *Pediatr. Clin.* **58**, 649–665 (2011).
56. Roenneberg, T., Allebrandt, K. V., Mero, M. & Vetter, C. Social jetlag and obesity. *Curr. Biol.* **22**, 939–943 (2012).

57. Gomez Fonseca, A. & Genzel, L. Sleep and academic performance: Considering amount, quality and timing. *Curr. Opin. Behav. Sci.* **33**, 65–71 (2020).
58. French, M. T., Homer, J. F., Popovici, I. & Robins, P. K. What you do in high school matters: High school GPA, educational attainment, and labor market earnings as a young adult. *East. Econ. J.* **41**, 370–386 (2015).
59. McLaughlin Crabtree, V. *et al.* Cultural influences on the bedtime behaviors of young children. *Sleep Med.* **6**, 319–324 (2005).
60. Guérin, N. *et al.* Role of school schedule, age, and parental socioeconomic status on sleep duration and sleepiness of parisian children. *Chronobiol. Int.* **18**, 1005–1017 (2001).
61. Edwards, F. Early to rise? The effect of daily start times on academic performance. *Econ. Educ. Rev.* **31**, 970–983 (2012).
62. Kim, T. The effects of school start time on educational outcomes: Evidence from the 9 O'Clock attendance policy in South Korea. *SSRN Electron. J.* **2019**, 1–26 (2018).
63. Shin, J. Sleep more, study less? The impact of delayed school start time on sleep and academic performance. (2018).
64. Jung, H. A late bird or a good bird? The effect of 9 o'clock attendance policy on student's achievement. *Asia Pacific Educ. Rev.* **19**, 511–529 (2018).
65. Lim, J. & Dinges, D. F. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychol. Bull.* **136**, 375 (2010).
66. Engle-Friedman, M. The effects of sleep loss on capacity and effort. *Sleep Sci.* **7**, 213–224 (2014).
67. Tarokh, L., Saletin, J. M. & Carskadon, M. A. Sleep in adolescence: Physiology, cognition and mental health. *Neurosci Biobehav. Rev.* **70**, 182–188 (2016).
68. Sadeh, A., Gruber, R. & Raviv, A. The effects of sleep restriction and extension on school-age children: What a difference an hour makes. *Child Dev.* **74**, 444–455 (2003).
69. Potkin, K. T. & Bunney, W. E. Sleep improves memory: The effect of sleep on long term memory in early adolescence. *PLoS ONE* **7**, 8–11 (2012).
70. Löwen, K., Baumert, J., Kunter, M., Krauss, S. & Brunner, M. Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project. In *The COACTIV Research Program: Methodological Framework* (eds Kunter, M. *et al.*) 79–96 (Springer, 2013).
71. Neumann, K., Kauertz, A. & Fischer, H. E. Quality of instruction in science education. In *Second International Handbook of Science Education* (eds Fraser, B. J. *et al.*) 247–258 (Springer, 2012). <https://doi.org/10.1007/978-1-4020-9041-7>.
72. Seidel, T. Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik* **60**, 850–866 (2014).
73. Seidel, T. & Shavelson, R. J. Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Rev. Educ. Res.* **77**, 454–499 (2007).
74. Hattie, J. Visible learning: A synthesis of over 800 meta-analyses relating to achievement. (2008). <https://doi.org/10.4324/9780203887332>.
75. Allen, J. D. Grades as valid measures of academic achievement of classroom learning. *Clear. House A J. Educ. Strateg. Issues Ideas* **78**, 218–223 (2005).
76. Heine, C., Briedis, K., Didi, H.-J., Haase, K. & Trost, G. *Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern* Vol. 3 (HIS-Kurzinformation, 2006).
77. Robbins, S. B. *et al.* Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychol. Bull.* **130**, 261 (2004).
78. Kaleli-Yilmaz, G. & Hanci, A. Examination of the 8th grade students' TIMSS mathematics success in terms of different variables. *Int. J. Math. Educ. Sci. Technol.* **47**, 674–695 (2016).
79. Estevan, I., Silva, A. & Tassinio, B. School start times matter, eveningness does not. *Chronobiol. Int.* **35**, 1753–1757 (2018).
80. Estevan, I., Silva, A., Vetter, C. & Tassinio, B. Short sleep duration and extremely delayed chronotypes in uruguayian youth: The role of school start times and social constraints. *J. Biol. Rhythms* 0748730420927601 (2020).
81. Arsintescu, L. *et al.* Validation of a touchscreen psychomotor vigilance task. *Accid. Anal. Prev.* **126**, 173–176 (2019).

Acknowledgements

We thank all the students for their engaged participation and the management of the Gymnasium Alsdorf, in particular W Bock and O Vollert, for their open support throughout the study. We also thank Dr. M Vuori-Brodowski for her contributions during wave 1, general advice and statistical support. This article has been published on bioRxiv with the following <https://doi.org/10.1101/2021.07.29.452310>.

Author contributions

Conceptualisation: E.C.W., T.R., Methodology: E.C.W., A.M.B., Investigation: A.M.B., E.C.W., C.M., Data curation: A.M.B., C.M., Formal analysis: A.M.B., E.C.W., F.O., G.Z., C.M., Validation: E.C.W., A.M.B., F.O., G.Z., Supervision: E.C.W., Visualization: A.M.B., E.C.W., Writing—original draft: A.M.B., E.C.W., C.F., Writing—review and editing: A.M.B., E.C.W., G.Z., F.O., C.F. All authors reviewed and approved the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. AMB received research and travel funds from the Graduate School of Systemic Neurosciences Munich.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-06804-5>.

Correspondence and requests for materials should be addressed to A.M.B. or E.C.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022