Later school start times in a flexible system improve teenage sleep

Eva C. Winnebeck[1,*], Maria T. Vuori-Brodowski[1], Anna M. Biller[1], Carmen Molenda[1], Dorothee Fischer[2,3,4], Giulia Zerbini[1] and Till Roenneberg[1,*]

[1]Institute of Medical Psychology, Ludwig Maximilian University Munich, Munich, Germany, [2]Department of Sleep and Human Factors Research, German Aerospace Center, Cologne, Germany, [3]Division of Sleep Medicine, Brigham and Women's Hospital, Boston, MA, and [4]Division of Sleep and Circadian Disorders, Harvard Medical School, Boston, MA

*Corresponding author: Eva C. Winnebeck, Institute for Medical Psychology, Goethestr. 31, 81373 Munich, Germany. Email: eva.winnebeck@med.uni-muenchen.de; Till Roenneberg. Email: till.roenneberg@med.uni-muenchen.de.

Abstract

Sleep deprivation in teenage students is pervasive and a public health concern, but evidence is accumulating that delaying school start times may be an effective countermeasure. Most studies so far assessed static changes in schools start time, using cross-sectional comparisons and one-off sleep measures. When a high school in Germany introduced flexible start times for their senior students—allowing them to choose daily between an 8 am or 9 am start (≥08:50)—we monitored students' sleep longitudinally using subjective and objective measures. Students (10–12th grade, 14–19 y) were followed 3 weeks prior and 6 weeks into the flexible system via daily sleep diaries ($n = 65$) and a subcohort via continuous wrist-actimetry ($n = 37$). Satisfaction and perceived cognitive outcomes were surveyed at study end. Comparisons between 8 am and ≥9 am-starts within the flexible system demonstrated that students slept 1.1 h longer when starting school later—independent of gender, grade, chronotype, and frequency of later starts; sleep offsets were delayed but, importantly, onsets remained unchanged. Sleep quality was increased and alarm-driven waking reduced. However, overall sleep duration in the flexible system was not extended compared to baseline—likely because students did not start later frequently enough. Nonetheless, students were highly satisfied with the flexible system and reported cognitive and sleep improvements. Therefore, flexible systems may present a viable alternative for implementing later school starts to improve teenage sleep if students can be encouraged to use the late-option frequently enough. Flexibility may increase acceptance of school start changes and speculatively even prevent delays in sleep onsets through occasional early starts.

Statement of Significance

In many cultures, teenagers are chronically sleep deprived because their typically late sleep times conflict with the relatively early start times of theirs schools. This is a pressing problem since teenage sleep deprivation is linked with reduced performance and substantial long-term health risks. However, the potentially simplest public countermeasure of delaying school starts requires more longitudinal, high-quality evidence. Our study adds important data on later school starts in Europe, using robust longitudinal comparisons and sleep measures, assessing a unique system of flexible start times. Sleep improved substantially and universally on days students opted to start classes at ≥9 am rather than 8 am. Net gains in the flexible system, however, required frequent late starts. Long-term effects of this system are under investigation.

Key words: sleep; adolescence; school start time; secondary school

Introduction

Adolescence is a decisive time in life, characterized by important developmental changes that shape individual future trajectories in health, education, social, and economic success. A recent review by Dahl and colleagues emphasized the importance of studying these modifications in order to develop policies to support adolescents during such a critical life period [1].

One marked—though often neglected—change during adolescence concerns sleep. Linked with pubertal development, adolescents show a progressive delay in the timing of their sleep until their early 20s when sleep time starts to advance again [e.g., 2–5]. Several biological, environmental, and social reasons have been suggested for explaining the later sleep in adolescents. First of all, the two biological processes regulating sleep—circadian and homeostatic—appear to be altered during adolescence [6]. The circadian system, which promotes wakefulness during the day and sleep at night, shows a later synchronization with the external day compared to children and adults [3, 7, 8] and thus provides a later circadian sleep window. At the same time, the build-up of sleep pressure appears slower, making adolescents less tired in the evening hours, which further delays their sleep [e.g., 9, 10]. This tendency for late sleep (not only on weekends but also throughout the school week) may be increased by external factors such as academic and peer pressure to stay up late studying or socializing online. Concomitantly, adolescents increase their exposure to evening light which results again in later sleep times [11, 12] by acutely increasing alertness [13–15] and potently delaying circadian rhythms [13, 16, 17]. The interplay between all these factors may thus result in a "vicious cycle of lateness" that exacerbates the natural (biological) tendency of sleeping late during adolescence.

Sleeping late *per se* would not be a problem if school schedules were organized accordingly. However, most schools have early start times that clash with adolescents' late sleep times. As a result, students accumulate a substantial lack of sleep over the school week [e.g. 18–22]. The consequences for performance and health are evident both in the short and long term. Negative effects of short sleep have been reported, among others, for academic performance [23], absenteeism and tardiness [24], participation and learning in class [25], emotional intelligence and constructive thinking skills [26], and motor vehicle accidents [e.g. 27, 28]. Even more worrying are the long-term health consequences of chronic sleep deprivation, such as increased risk for metabolic, cardiovascular, and inflammatory diseases [29, 30]; depressed mood [31–33]; and substance use [34, 35]. Additionally, students suffer from social jetlag, the mismatch between their circadian clock and their societal schedule [36]. Social jetlag, which is in most instances inherently coupled with sleep deprivation, has been linked with long-term health problems such as obesity and metabolic disorders [37–39].

An obvious solution to the problem of adolescent sleep deprivation is to delay school start times. Over the last decades, there has been much scientific effort to evaluate the impact of later start times. Most of the studies have been conducted in the US, and they have shown positive outcomes in terms of sleep duration and quality, mood, daytime sleepiness, concentration and attention in class, absenteeism, tardiness, and motor vehicle accidents [40–44]. Still, more studies are required not only in other countries to generalize the results but also to further substantiate the scientific evidence [45]. Given the school setting and research question, study designs are inherently limited and can thus usually not meet highest level evidence criteria such as randomization and double-blind placebo controls. However, so far, the majority of the designs has stopped short of what could be done by using cross-sectional rather than longitudinal comparisons. In addition, outcome parameters (e.g. sleep, mood, academic performance) have often been assessed with just a single-time questionnaire whereas longer monitoring, especially via objective measures such as activity recordings, are rare [42, 44, 46–50].

We had the opportunity to study the effects of later school starting times when a high school in Germany decided to introduce flexible start times for their senior students. Instead of fixed starts at mostly 8 am, in this new flexible system, the senior students could decide whether to start at 8:00 am or at 8:50 am (referred to as "9 am" herein for convenience) on a daily basis by attending or skipping the first period (a self-study period). We collected daily sleep data via diaries and, in >50% of participants, via objective, continuous activity measures over 9 weeks across systems and across early and late starts. This allowed us (1) to compare sleep between alternating early and late school starts within the flexible system in the same individuals without seasonal confounders and (2) to perform pre–post analyses in the same individuals to assess whether sleep changed from the rigid to the flexible system. This is one of the first studies assessing the effects of delayed school start times conducted in Europe and, to our knowledge, the first to assess the effects of flexible start times [51].

Methods

Study site

The study was performed at the Gymnasium Alsdorf, a high school in Alsdorf, Germany (50°53′N, 6°10′E). Alsdorf is a town of just below 50 000 residents situated in a former coal region in the very West of Germany. A gymnasium is the most academic of several types of secondary schools in the German educational system allowing access to higher education after successful completion. The Gymnasium Alsdorf received the German School Award in 2013 for its innovative teaching [52]. The school operates with a special educational concept called "Dalton plan," which includes daily self-study periods ("Dalton hours") for all students [53, 54]. During these self-study periods, students work through their personal 5-week curriculum with a teacher and on a subject of their choice. Each week, students had to fulfill a quota of 10 self-study periods.

School start times at baseline and in the flexible system

In order to address the late sleep times of their adolescent students, the school changed from a conventional school start system with fixed early start times to a new system with flexible start times (flexible system) for their senior students (10th–12th grade). In the conventional system, senior students started school at times predefined by their individual fortnightly schedules. This was usually at 8 am, a typical start time for German high schools, but included a later start on a median of 1 day a week (according to their schedules; cf. Figs. 1 and 6A).

With the introduction of the flexible system on February 1, 2016, one of the two daily self-study periods was moved into the first period (08:00–08:45), and senior students could decide on a daily basis whether to attend this first period or skip it and start school at 08:50 instead (referred to as "9 am" for convenience). Since some students' timetables included days (median ≈ 0.5 d/week) with a free period during the second period, skipping the first period on those days meant a school start at 10:15. Hence, we refer to all later starts as "≥9 am" to include also these cases.

Skipped self-study periods had to be fulfilled at another time during the week in one of the free periods in students' schedules. Although students usually had several free periods per week, there were individual limitations on how often the first self-study period could be skipped without getting home later than individual timetables would otherwise require (see example timetable Supplementary Table S1). Hardly any timetable allowed making up for five skipped self-study periods within its boundaries; however, no student would have had to stay later than the official 4.15 pm end to fulfill the weekly quota.

Study protocol

The recording period lasted from January 8 until March 14, 2016. We collected daily sleep diary data over 3 weeks before the transition to the flexible system and continued for another 6 weeks after the flexible system was introduced on February 1, 2016. We also collected objective sleep data via wrist-actimetry throughout the study period in a subcohort of students who also filled out daily sleep diaries. For a *status quo* assessment of sleep behavior at the beginning of the study,

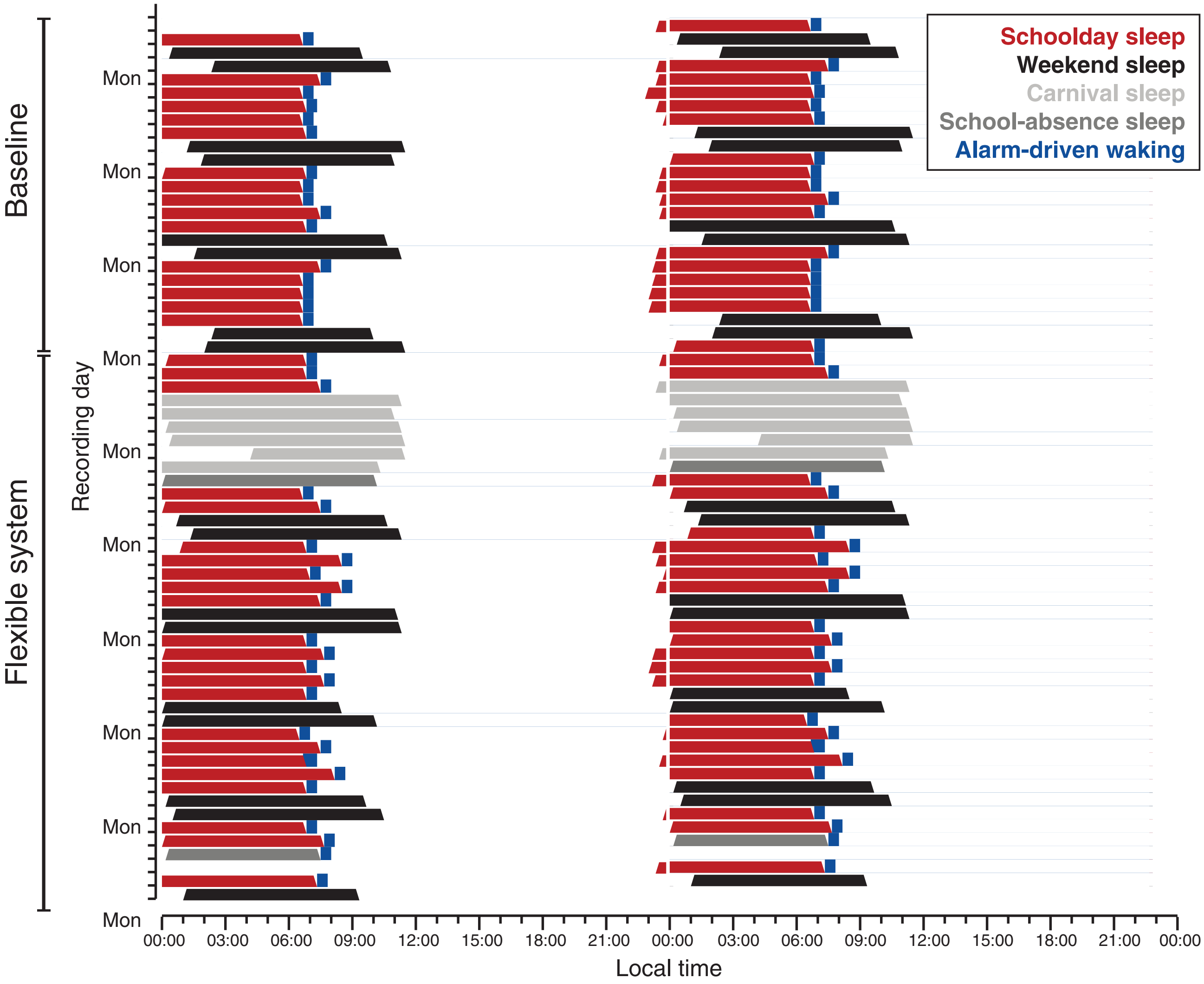

Figure 1. Sleep throughout the study period illustrating study design and nature of flexible system. Depicted are sleep-diary-recorded sleep episodes (colored bars) of one participating student over the entire study period. Data are double plotted. Data on nocturnal sleep episodes were collected over 9 weeks via an online sleep diary and simultaneously via actimetry in >50% of participants. During the first 3 weeks of recording (baseline), students started school at times predefined by their individual fortnightly schedules, which was usually at 8 am but included a later start at ≥9 am on around 1 day a week (median across full cohort; see red bars during baseline). Students were then followed 6 weeks into the new flexible system, where they could choose on a daily basis whether to attend the first period at 8 am or start school afterward at 9 am (08:50)—or occasionally even later on days if and when they had free periods afterwards (see red bars during the flexible system). The holiday period over carnival (light gray bars) was excluded from the analysis.

participants filled out the Munich ChronoType Questionnaire (MCTQ) [55, 56]; at the end of the study period, a purpose-designed survey about the flexible system was also filled out. The holiday period over carnival between February 4–9, 2016, was excluded from the analysis.

Participants

We informed all senior students and their parents or guardians via a study leaflet and orally during an information evening. All participants and at least one parent or guardian (when participant was < 18 y) had to provide written informed consent. The study was conducted in accordance with the Declaration of Helsinki and approved by the school board, the parent–teacher association and the student association of the school. Data were collected with a consent form that prohibits online deposition of data for open access sharing. This prohibition was implemented in order to protect participants' privacy in a cohort where most individuals are well-acquainted with each other and peers or teachers might identify participants.

We used opportunity sampling without specific exclusion criteria to maximize sample size. Of the 253 students attending 10th–12th grade (14–19 years) and thus transitioning into the flexible system, 113 (45%) signed up to participate in the study, 93 (82%) students provided at least some data, of which 65 (70%) passed our quantity and quality filter criteria for inclusion in the analysis. These criteria were (1) sleep information for ≥5 schooldays and ≥3 weekend days in each study phase (baseline and flexible system; 27 exclusions) and (2) congruent, plausible data (1 exclusion for reported wake-up times that were repeatedly in conflict with reported school start times). The final study cohort of 65 participants was used for all system comparisons. For comparisons between days with an 8 am or ≥9-am start, we additionally required sleep data from at least two 8 am-days and at least two ≥9 am-days per individual to ensure reliable comparisons. After applying this additional filter, a total of 60 participants remained in this subcohort. For activity recordings, teachers selected 45 students from all consenting participants who then additionally wore actimeters throughout the study period. After filter application, the actimetry subcohort consisted of 34 students also part of the diary cohorts. Cohort characteristics and sample sizes per participant are listed in Table 1.

Out of the 65 students from the full cohort, none reported use of any sleep medication, 3 students (5%) reported to be smokers, 12 (19%) reported weekly alcohol consumption of some

Table 1. Composition of study cohort and subcohorts.

		Cohort* Diary	Subcohort† Diary 8 am/≥9 am	Subcohort‡ Diary & Actimetry 8 am/≥9 am
Participants				
Total	*n*	65	60	34
Females	% (*n*)	62% (40)	63% (38)	65% (22)
Grade (10th/11th/12th)	% (*n*) per grade	40/35/25%	42/35/23%	32/38/29%
		(26/23/16)	(25/21/14)	(11/13/10)
Age (years)	Mean	16.5	16.5	16.7
	(SD, range)	(1.2, 14–19)	*(1.2, 14–19)*	(1.2, 14–19)
BMI	Mean	21.7	21.6	22.2
	(SD, range)	(2.9, 16.9–28.9)	(2.9, 16.9–28.9)	(3.0, 17.4–28.9)
Chronotype (local time)§	Mean	5.0	5.0	4.9
	(SD, range)	(1.0, 2.7–8.1)	(1.0, 2.7–8.1)	(0.88, 3.0–6.6)
Number of sleep diary entries per participant				
Baseline				
Days total	Median	21	21	22
(max. 24)	(IQR, range)	(20–23, 10–24)	(20–23, 10–24)	(21–23, 15–24)
Schooldays	Median	14(13–15, 6–16)	14	14
(max. 16)	(IQR, range)		(13–15, 6–16)	(13–15, 8–16)
Weekend days	Median	8	8	8
(max. 8 + absences)	(IQR, range)	(7–8, 3–8)	(7–8, 4–8)	(7–8, 5–8)
Flexible system				
Days total	Median	30	30	32
(max. 37)	(IQR, range)	(26–33, 9–37)	(26–33, 14–37)	(27–34, 16–37)
Schooldays	Median	20	20	21
(max. 27)	(IQR, range)	(16–22, 6–27)	(17–22, 9–27)	(19–23, 10–27)
Weekend days	Median	*10*	10	10
(max. 10 + absences)	(IQR, range)	(8–11, 3–15)	(8–11, 4–15)	(9–11, 4–15)
8 am-days	Median	11	11	11
	(IQR, range)	(8–16, 1–23)	(8–15, 2–21)	(8–15, 2–20)
≥9 am-days	Median	7	7	9
	(IQR, range)	(3–11, 0–19)	(4–11, 2–19)	(5–13, 2–19)

*Complete cohort (≥5 schooldays and ≥3 weekend days both at baseline and in flexible system).

†Subcohort for 8 am/≥9 am comparisons (additionally ≥2 days per start time in flexible system).

‡Subcohort for diary/actimetry comparisons (above filters also applied to actimetry data).

§MSFsc from MCTQ.

sort, and 49 (75%) reported weekly caffeine consumption, with caffeinated drinks as the main caffeine source—not tea or coffee (median of 0.6 drinks/day).

Munich Chronotype Questionnaire

At the beginning of the study all participants completed the Munich Chronotype Questionnaire (MCTQ) online [55–57]. We used a German version specifically designed for students where all questions pertaining to work were reworked to refer to school, and the formal German "you" (Sie) replaced with the informal "you" (Du) [57]. The MCTQ core module assesses sleep behavior on schooldays and school-free days, and additional modules pose questions about demographics, school times, commute to school, time spent outdoors, and substance use. An estimate of circadian phase of entrainment (chronotype) and a measure of circadian misalignment (social jetlag) are the core variables among the many variables obtainable from the MCTQ (see Data Analysis for formulae). Demographic data were taken from the MCTQ.

By definition, MCTQ-chronotype should only be interpreted if waking on free days is unrestricted, i.e. not alarm-driven; this was not the case for eight participants in the full cohort and for four participants in the subcohort 8 am/9 am. To avoid creating additional cohorts, we included these individuals in the analyses but established in sensitivity analyses without these individuals that results were essentially equivalent.

For comparisons of sleep behavior between our study cohort and other German adolescents, we randomly drew a 10-fold larger, age- and gender-matched sample of German adolescents from our MCTQ database on August 20, 2016. Because the study cohort contained three additional individuals at that time ($n = 68$ instead of 65; they were later eliminated during a last cleaning round), this database sample contains 680 individuals and not 650.

Sleep diary

To obtain daily records of participants' nocturnal sleep, we used a short online sleep diary based on the μMCTQ (a short version of the MCTQ) [58] adapted for a German student population. Students were asked to fill it out each morning throughout the study reporting on their past night's sleep. We sent reminder messages around twice a week. If students had missed to fill out the online diary for one or more instances, they were allowed to input their data at a later time point—in most of these instances, students reported to keep an offline log from which they then retrospectively populated the online diary. The sleep diary was provided via LimeSurvey.org. For further details on the diary itself and the data cleaning procedure, please refer to the extended methods in the SI.

Locomotor activity recording (actimetry)

Locomotor activity was recorded continuously over the entire study period in a subcohort of 45 participating students via wrist-worn activity-monitoring devices (Daqtometer, version 1.4, Daqtix, Germany). The data analysis pipeline via our in-house analysis program ChronoSapiens [59] entailed averaging activity counts per 30 s into 10-min-bins, excluding likely off-wrist periods (identified as stretches of 100 min of zero activity or as indicated in actimetry logs) and extracting estimated sleep bouts based on the identification of stretches of relative immobility as detailed in Roenneberg et al. [59]. To allow for sensible comparisons with diary recorded nocturnal sleep, daytime naps (any sleep occurring outside the daily 12-h-trough estimated via cosine fits [59]) were excluded, and bouts <180 min apart were combined into one longer bout. Please refer to the SI for more details.

Final survey

We developed a 12-item self-assessment questionnaire to obtain additional information about the individual use of and satisfaction with the flexible system and the perceived cognitive outcomes. This survey was completed by 56 of the full cohort of 65 students and anonymously by another 82 senior students in the flexible system to assess any selection bias. The participants received the paper-pencil survey in German on the last day of data collection and completed it immediately.

The first six items examined the use of the flexible system. The students were asked to indicate (1) whether they were satisfied with the new system (yes/no), (2) whether it was difficult for them to start school at 8 am (never/mostly/always), (3) whether it was easier to start school at 9 am compared to 8 am (never/mostly/always), (4) how often (0 days/1–2 days/3–4 days/5 days) and (5) on which days of the week they attended the first period at 8AM (Mo/Tu/We/Th/Fr), and (6) reasons for starting school at 8 am. Here, they were given the possibility to state their own reasons or cross at least one of eight alternatives (easier to study/easier to get to school/additional study time/friends/specific self-study teacher/specific subject/fulfill self-study quota/other).

The final six items assessed the behavior and feeling of the students during baseline and the flexible system. The first item asked about sleep duration in hours and the second about alarm-driven waking (0–5 days). The last four items assessed the quality of sleep, how tired the students felt, ability to concentrate in class, and ability to study at home after school. Each item was scored on a five-point Likert scale (1 = "bad/poor" to 5 = "good").

Data analysis

Analyses and visualization were performed in SPSS Statistics (IBM, version 24 and 25) and R [60] (versions 3.5.1 "Feather Spray" and 3.5.3 "Great Truth") using the R packages effsize [61], ggplot2 [62], ggpubr [63], Hmisc [64], lmer4 [65], lmerTest [66], PMCMRplus [67], RColorBrewer [68], and reshape2 [69].

Data aggregation

For analyses, time course data were aggregated via mean (median for the ordinal variable *sleep quality rating*) to one data point per individual for the six conditions of interest. These conditions were (1) baseline schooldays, (2) baseline weekends, (3) flexible system schooldays, (4) flexible system weekends, (5) flexible system 8 am-days, and (6) flexible system ≥9 am-days. Over the carnival holidays during the flexible system (February 5–9, 2016), students' diary compliance was reduced. The remaining entries indicated more irregular sleep, delayed sleep timing and daytime sleep. To minimize any influence on results, we excluded the carnival period from the free-day aggregates, which are based on fewer data points and can thus be more easily distorted by outliers (Table 1). However, we included the schoolday sleep following the holidays in the schoolday-aggregate measures (as examples of schoolday sleep after a party weekend), where potential outliers are balanced out by more data points (Table 1).

Derived data

From the aggregated measures, the following variables were calculated as per the equations below: average daily sleep duration across the week (SD_{week}); midsleep on schooldays (MSW); midsleep on school-free days (MSF); chronotype as MSF corrected for oversleep (MSF_{sc}); social jetlag (SJL); difference and ratio between ≥9 am-days and 8 am-days for variables of interest ($DELTA_x$; $RATIO_x$, respectively); frequency of ≥9 am-starts (also referred to as 9 am-use) and of alarm-driven waking.

$$\mathrm{SD_{week}} = \frac{\mathrm{SD_{schooldays}} \times 5 + \mathrm{SD_{freedays}} \times 2}{7}$$

$$\mathrm{MSW} = \mathrm{Sleep\ Onset_{schooldays}} + \frac{1}{2}\mathrm{SD_{schooldays}}$$

$$\mathrm{MSF} = \mathrm{Sleep\ Onset_{free\ days}} + \frac{1}{2}\mathrm{SD_{free\ days}}$$

$$\mathrm{MSF_{SC}} = \mathrm{Sleep\ Onset_{free\ days}} + \frac{1}{2}\mathrm{SD_{week}}$$

$$\mathrm{SJL} = \mathrm{MSF} - \mathrm{MSW}$$

$$\mathrm{DELTA}_x = x_{9\,\mathrm{am\text{-}days}} - x_{8\,\mathrm{am\text{-}days}}$$

$$\mathrm{RATIO}_x = \frac{x_{9\,\mathrm{am\text{-}days}}}{x_{8\,\mathrm{am\text{-}days}}}$$

$$\text{frequency of} \geq 9\,\mathrm{am} - \mathrm{starts} = \frac{n_{9\,\mathrm{am\text{-}starts_{flex}}}}{n_{\mathrm{schoolday\text{-}entries_{flex}}}} \times 100$$

$$\text{frequency of alarm-driven waking} = \frac{n_{\mathrm{alarm-driven\ waking_{flex}}}}{n_{\mathrm{schoolday-entries_{flex}}}} \times 100$$

Statistical analysis

Data analysis was in part hypothesis driven (comparisons between 8 am/≥9 am-days and between systems) and in part exploratory (analysis of benefit and 9 am-use) to identify important unpredicted patterns. All statistical tests were evaluated to a significance level of $\alpha < 0.05$ based on two-sided tests. We used parametric tests for all analyses unless data were below interval level or Shapiro-Wilk test indicated nonnormal distribution of a variable in at least one group.

Unfortunately, we could not combine analyses of baseline/flexible system and 8 am/≥9 am-starts since we lacked reliable information on exact school start times during baseline for each day and participant. We had not asked students about their daily school start time during the baseline period, and students did not follow their timetables exactly (due to teacher absences, exams, etc.; identified via clear mismatches between timetable information and reported wake-up times). Hence, we performed separate analyses as detailed below.

For comparison of sleep parameters between 8 am and ≥9 am-days in the flexible system, we performed paired *t*-tests or Wilcoxon signed-rank tests. Effect size was subsequently estimated using either Cohen's *d* after paired *t*-tests via the R package effsize [61] or using the procedure described by Rosenthal [70] for Wilcoxon signed-rank tests ($r = Z/\mathrm{sqrt}(N_{observations})$; $N_{observation}$ was 2*cohort size as data were paired).

For comparison of sleep parameters between baseline and the flexible system, we used two approaches. For variables present for both schooldays and weekends, we performed 2-factorial repeated-measures ANOVAs with system (baseline/flexible system) and weekday (schoolday/weekend) as main effects. When interaction effects system × weekday were statistically significant, we performed *post hoc* pairwise comparisons via *t*-tests testing for differences between baseline and flexible system. With two *t*-tests performed per variable, we corrected *p*-values via the Bonferroni method by multiplication by 2 to control the familywise error rate. For variables incorporating information from both schooldays and weekends (social jetlag and daily mean sleep duration across week), we performed paired t-tests or Wilcoxon signed-rank tests as described above.

Frequency of alarm-driven waking was additionally analyzed via logistic regression because of the large ceiling effect in this variable (cf. Figures 2H and 4D). To this end, frequency of alarm-driven waking was dichotomized into high and low frequency of alarm-driven waking based on a median split: at 100% alarm-driven waking for 8 am versus ≥9 am-days; at 93% for baseline versus flexible system. Results were equivalent in their direction and statistical significance when using two other splits: (1) split at 1st quartile (85% alarm-driven waking) and (2) discontinuous split below 1st quartile versus 100% (Supplementary Tables S2 and S3). Logistic regression was performed via mixed effects models using the R package lme4 [65] to accommodate the repeated measures nature of the data by including ID as a random effect. In the models reported here, we also included gender as covariate, since there was an obvious trend that males were woken more often by an alarm than females in the flexible system. However, neither exclusion of the covariate gender nor inclusion of additional covariates such as age, chronotype (MCTQ-MSF_{sc}), or 9 am-use altered the effect of school start time or school system in a notable way. Also, gender never reached statistical significance at $p < 0.05$ in any of the models.

For the exploratory analysis of characteristics associated with a benefit (sleep extension) and 9 am-use, we analyzed data via Pearson or Spearman correlations for continuous variables as well as via unpaired *t*-tests, Wilcoxon rank sum tests, one-way ANOVA, or Kruskal–Wallis tests for group comparisons.

The correspondence between diary-recorded and actimetry-determined sleep was assessed via Pearson correlations for average sleep onsets or sleep offsets per person and the 5 relevant, nonoverlapping conditions (see data aggregation) leaving out the 6th condition "flexible system schooldays." Differences between the full study cohort ($n = 65$) and the age- and gender-matched MCTQ-database sample were assessed via Wilcoxon rank sum tests.

Results of statistical tests are reported in the main text in brackets, listing the specific test statistic, the p value and, if applicable, the effect size. Where results across similar variables with similar outcomes are provided in the same bracket, we listed the ranges of the above values across variables. The tests statistics indicate the following tests: t, t-test; W, Wilcoxon rank sum test; Z, Wilcoxon signed-rank test; r, Pearson correlation; rho, Spearman correlation; H, Kruskal–Wallis test.

Results

Study cohort

The total study cohort, after exclusions based on minimum quantity and quality criteria for sleep diary entries, comprised 65 adolescent students aged 14–19 years covering all three school grades that transitioned to the flexible system (Table 1). The median record length per participant was 21 nocturnal sleep episodes in baseline and 30 episodes in the flexible system. Depending on the study question, we also used two subcohorts for analyses, both of which were very similar in their characteristics to the main cohort (Table 1).

Sleep of study cohort is similar to that of other German adolescents

To determine how representative the sleep of our participating students was of other German adolescents, we compared key sleep parameters of the study cohort, assessed via the MCTQ at the beginning of the study, to a ~10-fold larger, age- and gender-matched German sample ($n = 680$) from our large MCTQ database. Study participants were indistinguishable from the larger database sample in any of the analyzed parameters. Namely, sleep duration on schooldays and school-free days, chronotype

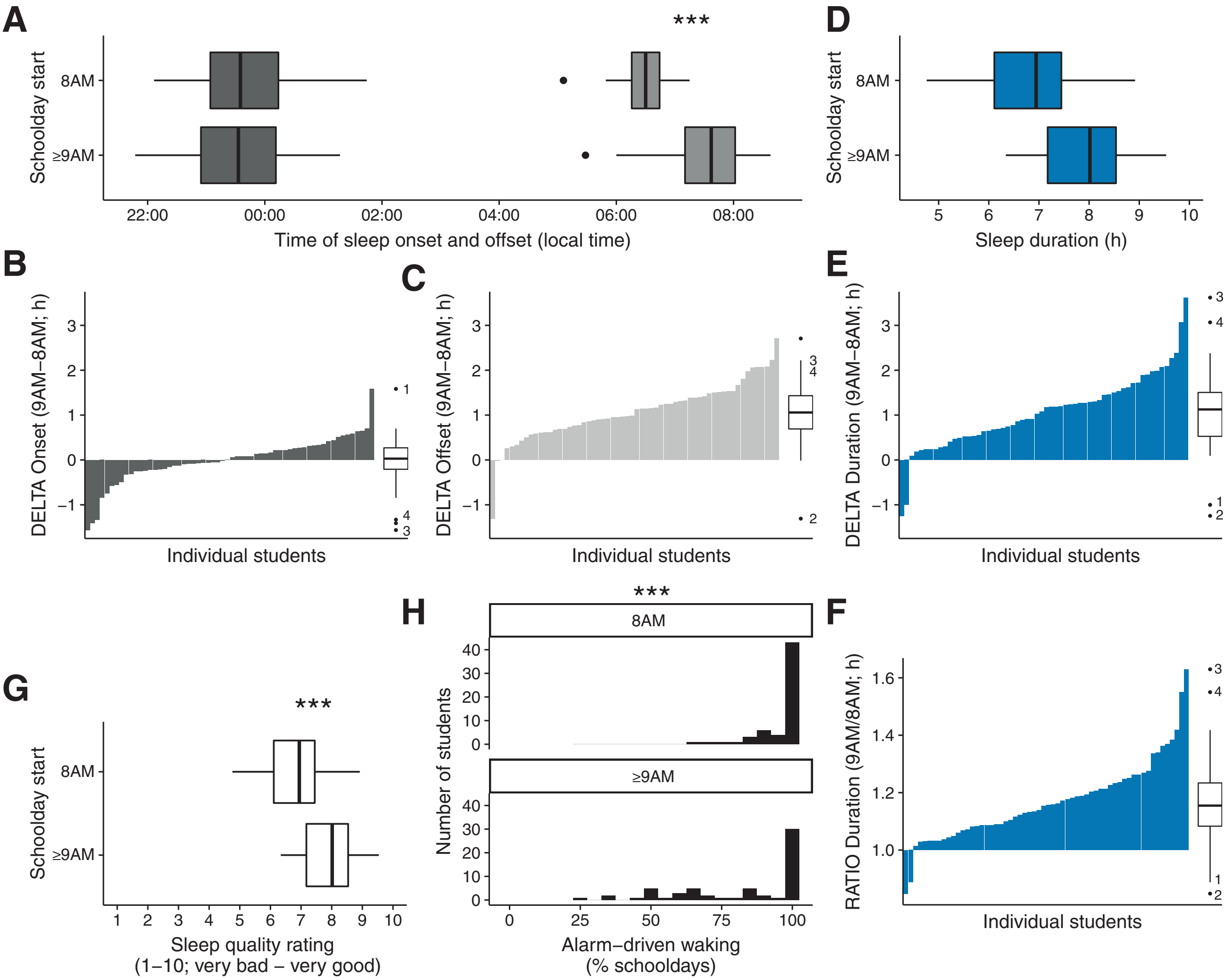

Figure 2. Comparison of sleep parameters between 8 am-days and ≥9 am-days in the flexible system. Sleep parameters are from sleep diaries of the subcohort for 8 am/9 am-comparison ($n = 60$). (A) Average sleep onset (dark gray) and offset (light gray) times on 8 am and ≥9 am-days. The average absolute difference in these measures for each individual is depicted in (B) for sleep onset times (DELTA Onset) and in (C) for sleep offset times (DELTA Offset). (D) Average sleep duration on 8 am and ≥9 am-days. Each individual's average difference in sleep duration is depicted in (E) in absolute terms (DELTA Duration) and in (F) in relative terms (RATIO Duration). (G) Average sleep quality rating. (H) Distributions of individuals' frequency of alarm-driven waking on schooldays. Statistical analysis was performed via paired t-tests or Wilcoxon signed-rank tests with $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; Tukey boxplots. Numbers 1–4 identify the 4 over- and underbenefitting students as per DELTA duration in E.

(midsleep on free days; MSF_{sc}) and social jetlag appeared the same (range of W: 19052–24558; range of p: 0.066–0.2592; Supplementary Figure S1). Furthermore, study participants also displayed the gender difference in MCTQ-derived chronotype common for this age group [2], with chronotype on average 1.1 h later in male than female participants ($t(48) = 4.628$; $p < 0.0001$; $d = 1.202$), altogether indicating that our sample shows sleep behavior typical for German adolescents—late sleep timing, short sleep on schooldays, long sleep on school-free days, and high social jetlag.

Self-reported sleep times match objective sleep data

Based on data from the subcohort of participants wearing actimeters and filling out sleep diaries simultaneously ($n = 34$), we found that subjective, self-reported sleep times matched well with objective sleep times determined from the actimetry records. Average sleep onsets and offsets from both measures were highly correlated ($r = 0.91$ and $r = 0.94$; $p < 0.0001$) and also essentially equivalent (Supplementary Figure S2), indicating that the cohort faithfully reported sleep times. Based on this validation, we opted for an analysis of the larger cohort with sleep diary data rather than focusing on the smaller actimetry subcohort.

Sleep on 8 am versus ≥9 am-days in the flexible system

This section presents our analyses of students' sleep *within* the flexible system. Here, we compared average sleep on nights before a normal 8 am school start (8 am-days) to that on nights when students took advantage of the new option to skip the first period and started school at 9 am—or occasionally even later if they had additional free period(s) in their individual timetables afterward (≥9 am-days). For simplicity, we henceforth speak of "sleep *on* 8 am-days" or "*on* ≥9 am-days" and mean this to be the nocturnal sleep episodes preceding days with an 8 am or ≥9 am-school start.

Frequency of ≥9 am-starts

Students varied substantially in their use of the 9 am-option, which ranged from 0% to 90% of a student's recorded schooldays (cf. Figure 6). The median frequency of ≥9 am-starts was 39% (IQR: 20%–60%), which amounts to 2 days out of a 5-day school week. For the following analyses, only students that used both the 8 am-option and the 9 am-option at least twice were included ($n = 60$, subcohort 8 am/9 am, Table 1).

Sleep onset, offset, and duration

As expected, on ≥9 am-days, students woke later than on 8 am-days ($t(59)=–13.017$; $p < 0.0001$; $d = 1.68$; Figure 2A). The mean difference in their sleep offset was 1.1 h (SD: 0.64 h)—a larger difference than anticipated for a 50-min delay in school start. There are two likely and additive reasons for this large effect: (1) almost every single student delayed his/her sleep offset time on ≥9 am-days (DELTA Offset >0 h, Figure 2C), and (2) several students had an additional free period after the skipped first period on several of their ≥9 am-days, allowing them to delay their wake-up times far beyond the expected 50-min difference (DELTA Offset >> 0.83 h, Figure 2C). Both communication with the school and our retrospective checks of students' timetables confirmed that this was the case.

Importantly, despite their later sleep offset times, students did not systematically delay their sleep onset times on ≥9 am-days ($t(59)=0.0259$; $p = 0.9794$; $d = 0.003$; Fig. 2A), illustrated by an even number of students falling asleep either slightly earlier or later on ≥9AM-days compared to 8AM-days (DELTA Onset, Fig. 2B).

Given these stable sleep onsets and markedly delayed offsets, sleep duration was longer on ≥9AM-days than on 8AM-days ($Z = 6.27$, $p < 0.0001$, $r = 0.57$; Fig. 2D). Students extended their sleep by 1.1 h (median; IQR: 0.53–1.5 h) or 15% (median; IQR: 8%–23%) from a median of 6.9 h to 8.0 h on ≥9 am-days (Figure 2D, E and F). Again, the great magnitude of the effect likely results from the 9 am-option sometimes representing a ≥9 am-option as well as almost all students extending their sleep on ≥9 am-days (Figure 2E and F).

Subjective sleep quality

On ≥9 am-days, students rated their sleep quality higher than on 8 am-days ($Z = –4.435$, $p < 0.0001$, $r = 0.40$; Figure 2G). The median increase was 0.8 points (IQR: 0–1.6) on a 10-point rating scale.

Alarm-driven waking

The proportion of schooldays on which students indicated "woken by alarm clock" was substantial: all students were woken by their alarm more than once a week, and half of the students reported alarm-driven waking on all of their schooldays on both 8 am and ≥9 am-days (Figure 2H; median in both conditions = 100% of schooldays). Because of this marked ceiling effect in alarm-driven waking, analyses may be less reliable, so we used not only a nonparametric test but also logistic regression to assess potential differences between 8 am and ≥9 am-days. Both analyses indicated that, although the rate of alarm-driven waking was still high on ≥9 am-days, students were woken less often by their alarm than on 8 am-days ($Z = 4.55$, $p < 0.0001$, $r = 0.42$), and the odds for less alarm-driven waking (<100% of schooldays, i.e. alarm-free waking on several schooldays) were increased on ≥9 am-days (OR = 3.3; 95% CI = 1.28–8.48; Supplementary Table S2).

Extension of sleep on ≥9 am-days was independent of gender, grade, chronotype, and frequency of ≥9 am-starts

To understand which type of student may particularly benefit from later starts, we searched for factors linked with sleep extension on ≥9 am-days, which we considered the core measurable benefit in our study. Sleep extension was quantified as each student's difference in sleep duration between their ≥9 am and 8 am-days, in either absolute terms (DELTA sleep duration 9 am–8 am; Figure 2E) or relative terms (RATIO sleep duration 9 am/8 am; Figure 2F). Below, only the results for absolute sleep extension are presented since results for relative sleep extension were essentially equivalent.

The amount of sleep extension on ≥9 am-days showed no systematic relationships with any of the "key suspects" that we assessed. There was no evidence that genders benefitted differently ($t(57.3) = –0.2109$; $p = 0.8337$; $d = –0.0711$; Figure 3A) or that students from a certain grade (implicitly incorporating the factor age) benefitted more or less ($H(2)=2.6445$; $p = 0.2665$, Figure 3B). Notably, also chronotype (either MCTQ or sleep-diary-derived MSF_{sc} at baseline or flexible system) was not associated with the amount of sleep extension (range of r: –0.22–0.06; range of p: 0.0845–0.6234; Figure 3C).

The apparent lack of influence of any of the above factors tallies with the fact that the benefit from ≥9 am-starts was close to universal: virtually all participating students (97%, 58 out of 60) slept longer on ≥9 am-days than on 8 am-days (Figure 2E). There were only two students that did not benefit (DELTA sleep duration <0 h; outliers 1 and 2 in Figure 2E), contrasting with two students who benefitted over-proportionally (DELTA sleep duration >3 h; outliers 3 and 4 in Figure 2E).

What stood out for these negative and positive outliers in sleep extension (Figure 2E) was that they were at opposite ends in their 9 am-use: the two overbenefiters rarely made use of the 9 am-option, whereas the two non-benefiters started quite often at ≥9 am (Figure 3D). This could have indicated that going more often at ≥9 am reduces the benefit from late starts—a potentially central problem invalidating the flexible system. Indeed, at first sight, this was supported by a negative correlation between 9 am-use and DELTA sleep duration across all students (ρ = 0.33, p = 0.0112; Figure 3D). However, this association was only driven by exactly these four outliers. When excluding these from the analysis, the amount of benefit is not associated with the frequency of ≥9 am-starts anymore (ρ = −0.22, p = 0.1060; Figure 3D). Furthermore, the most likely mechanism for a smaller sleep extension with greater 9 am-use would be a delay in sleep onsets on ≥9 am-days. However, there is no hint that students with greater 9 am-use had relatively delayed sleep onsets on ≥9 am-days since DELTA onset was not correlated with the frequency of ≥9 am-starts (ρ = 0.05, p = 0.7078; Figure 3E). However, DELTA offset shows such a correlation (ρ = 0.30, p = 0.0237; Figure 3F): students with the least 9 am-use had the greatest delay in offsets. This cross-check shows clearly that a substantial proportion of students with low 9 am-use benefitted overproportionally through delaying their offsets far beyond the 50-min-extension—likely by starting school much later than 9 am on the few days that they skipped the first period—in contrast to the high users who regularly went at truly 9 am. Hence, the data provide no indication that the frequency of later starts systematically affected the benefit.

Sleep in the flexible system versus baseline

Surprisingly, the switch to the flexible system did not markedly improve students' sleep: Most sleep parameters in the flexible system were not or only minimally different from those reported during the baseline period with fixed school start times (Figure 4).

Sleep onset, offset, and duration

At first glance, the results are perfectly in line with the positive expectations elicited by the above results comparing sleep on 8 am and ≥9 am-days in the flexible system. Average sleep onset times on schooldays were the same between baseline and flexible system ($t(64) = -0.764$; $p_{\mathrm{bonf}} = 0.8956$; *post hoc* test to two-way ANOVA as reported in Figure 4A), whereas sleep offset times on schooldays were delayed in the flexible system ($t(64) = 2.496$; $p_{\mathrm{bonf}} = 0.0303$; *post hoc* test Figure 4A). However, this delay sports only a small statistical effect size ($d = -0.205$) and is small also in biological terms at only 6 min (SD: 24 min). Accordingly, average

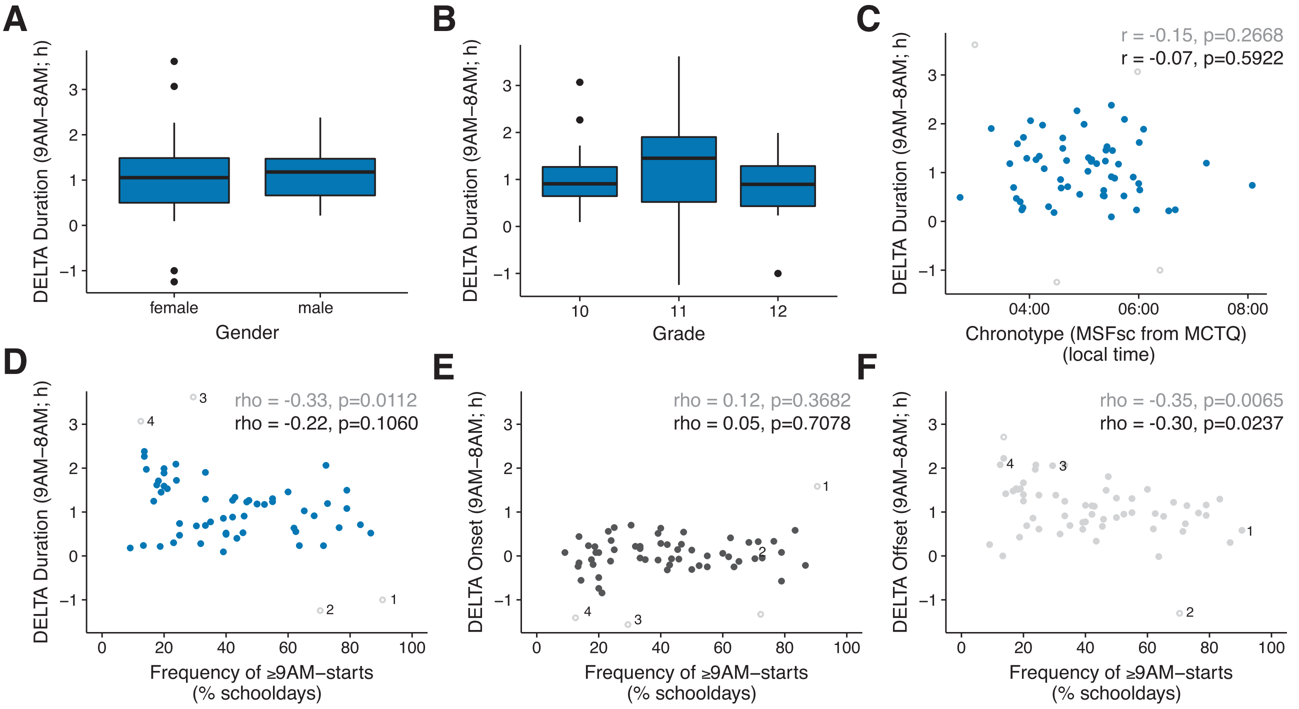

Figure 3. Extension of sleep on ≥9 am-days in the flexible system appears independent of gender, grade, chronotype and frequency of ≥9 am-starts. Depicted are the absolute differences (DELTA values) in sleep parameters between ≥9 am and 8 am-days in the flexible system and their relationship to other variables. Sleep parameters are from sleep diaries of the subcohort for 8 am/9 am-comparisons ($n = 60$). (A,B,C) show the difference in sleep duration between ≥9 am and 8 am-days (sleep extension) against (A) gender, (B) grade, (C) chronotype (midsleep on school-free days corrected for oversleep). (D,E,F) show the relationship between frequency of ≥9 am-starts (percentage of schooldays that a student started school at ≥9 am) and the difference between ≥9 am and 8 am-days in (D) sleep duration (DELTA Duration = sleep extension), (E) in sleep onset (DELTA Onset) and (F) in sleep offset (DELTA Offset). Data are color-coded as in Figure 2 and numbers 1–4 identify the same 4 over- and underbenefitting students. Tukey outliers in the y-axis variable are marked by gray empty circles. Results of Pearson and Spearman correlations are given for data both including outliers (gray) and excluding outliers (black). Statistical analysis for A was via unpaired t-test and for B via Kruskal-Wallis test.

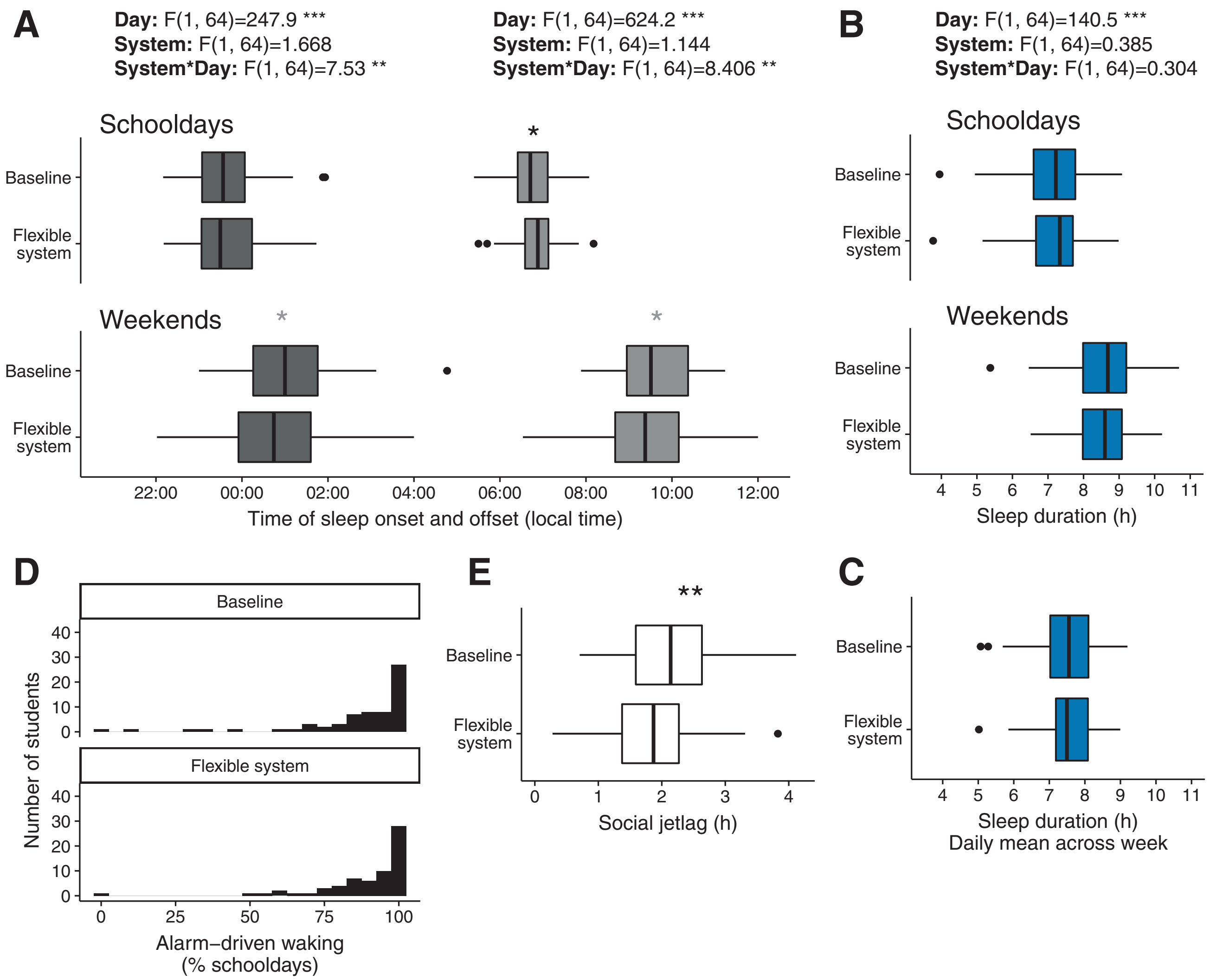

Figure 4. Comparison of sleep parameters between baseline and the flexible system. Sleep parameters are from sleep diaries of the full cohort ($n = 65$). (A) Average sleep onset (dark gray) and offset (light gray) times, and (B) average sleep duration for both study phases on schooldays and weekends. Results of two-way ANOVAs for the factors weekday (schoolday/weekend) and system (baseline/flexible system) are given above the respective graphs. $^*p < 0.05$ indicates results of *post hoc* tests before (gray) and after (black) Bonferroni correction. (C) Average mean daily sleep duration across a week (weighted for 5 schooldays and 2 weekend days). (D) Distributions of individuals' frequency of alarm-driven waking on schooldays. (E) Average social jetlag at baseline and in the flexible system. Statistical analysis was performed via paired *t*-tests or Wilcoxon signed-rank tests with $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; Tukey boxplots.

sleep duration on schooldays was indistinguishable between both school start systems ($t(64) = -1.100$; $p_{\text{bonf}} = 0.551$; *post hoc* test), just like weekend sleep duration and the daily mean duration across the entire week (all $p_{\text{bonf}} \geq 0.3649$; Figure 4B and C).

Alarm-driven waking

The proportion of schooldays on which students were woken by their alarm was as high in the flexible system (median: 93%; IQR: 86%–100%) as during baseline (median: 95%; IQR: 85%–100%; $Z = 0.47$, $p = 0.64$, $r = 0.04$). The odds ratio from logistic regression also did not indicate a change in odds for alarm-driven waking between school systems (OR = 0.69; 95% CI = 0.29–1.62; Supplementary Table S3) (Figure 4D).

Social jetlag

Overall, the typically large differences in sleep timing and duration between schooldays and weekends were found both during baseline and the flexible system (Figure 4A and B). However, sleep timing on weekends became slightly earlier in the flexible system, with both onset and offset advanced by a mean of 12 min (SD: 54 min and 48 min)—a difference with small effect size and only statistically significant before correction for multiple testing (onset: $t(64) = 2.092$; $p_{\text{bonf}} = 0.0808$; $d = 0.200$; offset: $t(64) = 2.264$; $p_{\text{bonf}} = 0.0539$; $d = 0.227$; *post hoc* tests). In combination with the later offset times on schooldays, however, this trend toward earlier sleep on the weekend resulted in a slight reduction in students' social jetlag by 18 min (SD: 42 min) in the flexible system ($t(64) = 3.309$; $p = 0.0015$; $d = 0.411$) (Figure 4E). Although this suggests a positive effect of the flexible system, we urge for a cautious interpretation, given the small effect sizes and the manifold unsystematic reasons why sleep timing might have become earlier on the weekend days monitored. Furthermore, a systematic advance of sleep timing in spring following the advance of dawn has been found in several studies [71, 72] and could explain this effect.

Subjective improvements in the flexible system

Interestingly, although the daily sleep diary entries did not indicate a general improvement in sleep parameters between baseline and flexible system, students nonetheless felt that they were faring better overall in the new system (Figure 5). In our survey at the end of the study, which was filled out by 56 of the 65 participants, students estimated their sleep times to be 0.5 h longer (median) in the flexible system than at baseline ($Z = 5.15$, $p < 0.0001$, $r = 0.49$) and also rated their sleep quality higher ($Z = 4.83$, $p < 0.0001$, $r = 0.46$) (Figure 5A and B). Merely their alarm need was not altered in their view ($Z = 1.36$, $p = 0.17$, $r = 0.13$; Figure 5C). In terms of cognitive improvements, students felt that they were less tired ($Z = 4.67$, $p < 0.0001$, $r = 0.44$) and could concentrate better during class ($Z = 5.07$, $p < 0.0001$, $r = 0.48$) and that their ability to study at home after school was improved ($Z = 3.88$, $p = 0.0001$, $r = 0.37$) (Figure 5D and F).

Frequent ≥9 am-starts in the flexible system are associated with longer sleep in the flexible system

The discrepancy of a universal sleep benefit from ≥9 am-starts but not obviously from the flexible system overall might result from students' low use of the 9 am-option in the flexible system.

Frequency of ≥9 am-school starts in the flexible system versus the conventional system

In the flexible system, students used the 9 am-option on average on only 2 days per week (median: 39%, IQR: 20%–60% of schooldays,

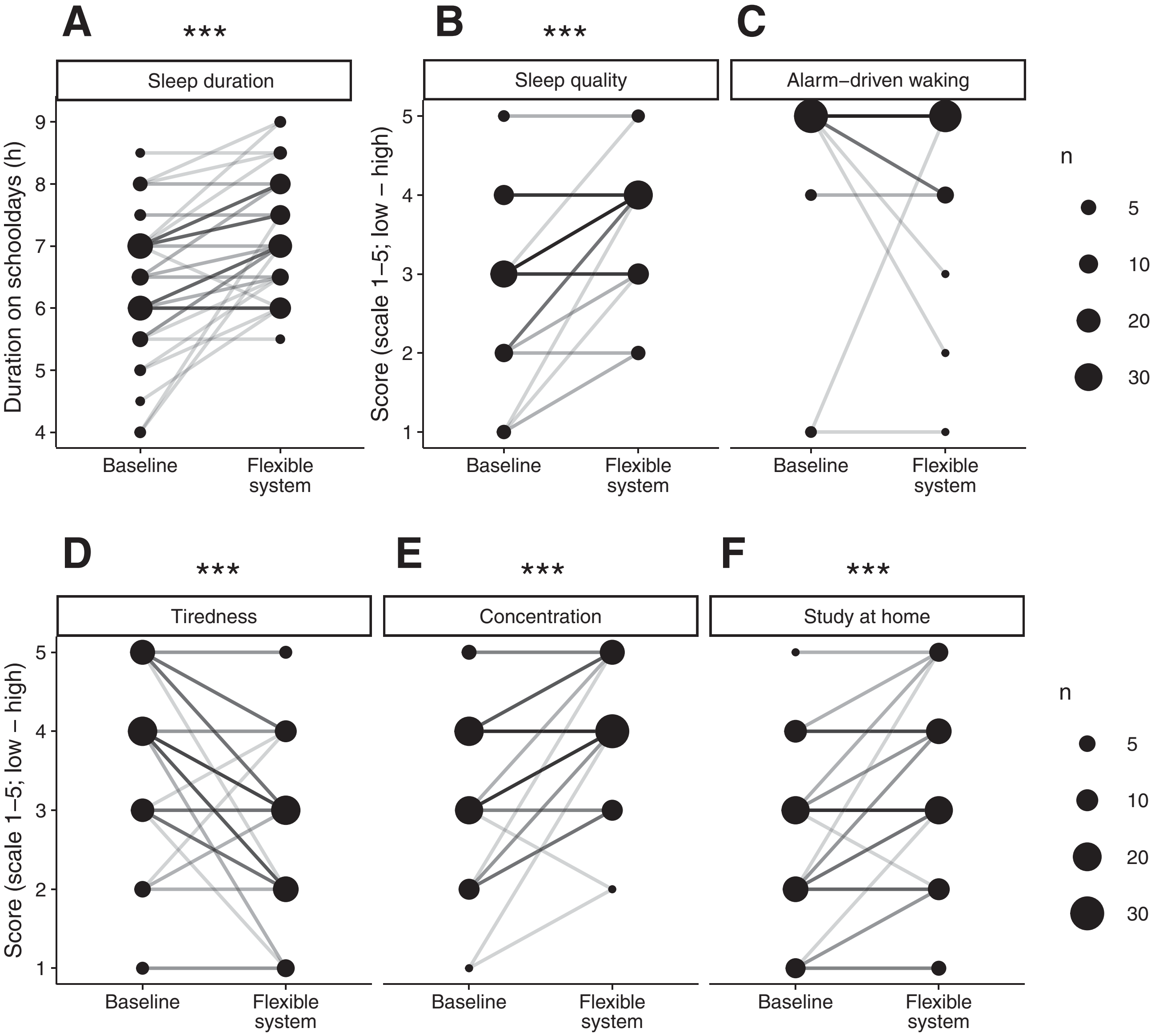

Figure 5. **Student-rated benefit from the flexible system.** Depicted are the results of the final survey in which participants rated their subjective experience in the conventional and flexible system. Data are from 56 students of the main cohort of 65 (9 students did not return the final survey). (A) Sleep duration on schooldays in hours. (B) Sleep quality on schooldays. (C) Frequency of alarm-driven waking per school week. Subjective score for (D) tiredness during class, (E) ability to concentrate during class, and (F) ability to study at home after school. Data are displayed in bubble charts to represent the categorical nature of the data. The area of each circle indicates the number of data points represented. Lines show trajectories of each individual (within-subject trajectories); the darkness of each line illustrates the number of individual trajectories it represents. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ for Wilcoxon signed-rank tests.

Figure 6A). The frequency of 9 am-use was not stable across the 6 weeks monitored but tended toward highest values in week 3 and lowest in week 6 (see Supplementary Figure S3). In the conventional system at baseline, students had no scheduled first period on ~1 day/week (median: 20%, IQR: 14%–27% of schooldays, Figure 6A), a median difference of only 0.75 days from the flexible system (IQR: 0.2–1.7 days; $Z = 5.35$, $p < 0.0001$, $r = 0.47$). This small increase in the number of later starts in the flexible system might hence be the reason for the lack in measurable sleep benefit in our study.

Frequency of ≥9 am-starts in the flexible system—associated with sleep duration but not sleep timing

We therefore sought to determine whether greater use of the 9 am-option was linked with better sleep in the flexible system. Although we could not perform direct comparisons *between* the systems factoring in start times (due to lack of information about exact start times during each day in baseline, see methods), we were able to check for associations *within* the flexible system.

Above, we demonstrated that the benefit of going to school at ≥9 am (DELTA sleep duration) was not affected by how often students actually went at ≥9 am—it was similarly high for all 9 am-use frequencies (Figure 3D). Our broader analyses here—looking at sleep parameters *across* the flexible system instead of ≥9 am to 8 am-day differences—show that making greater use of the 9 am-option is clearly associated with longer sleep in the flexible system. The more frequently students started school later, the longer was both their average schoolday sleep ($\rho = 0.39$;

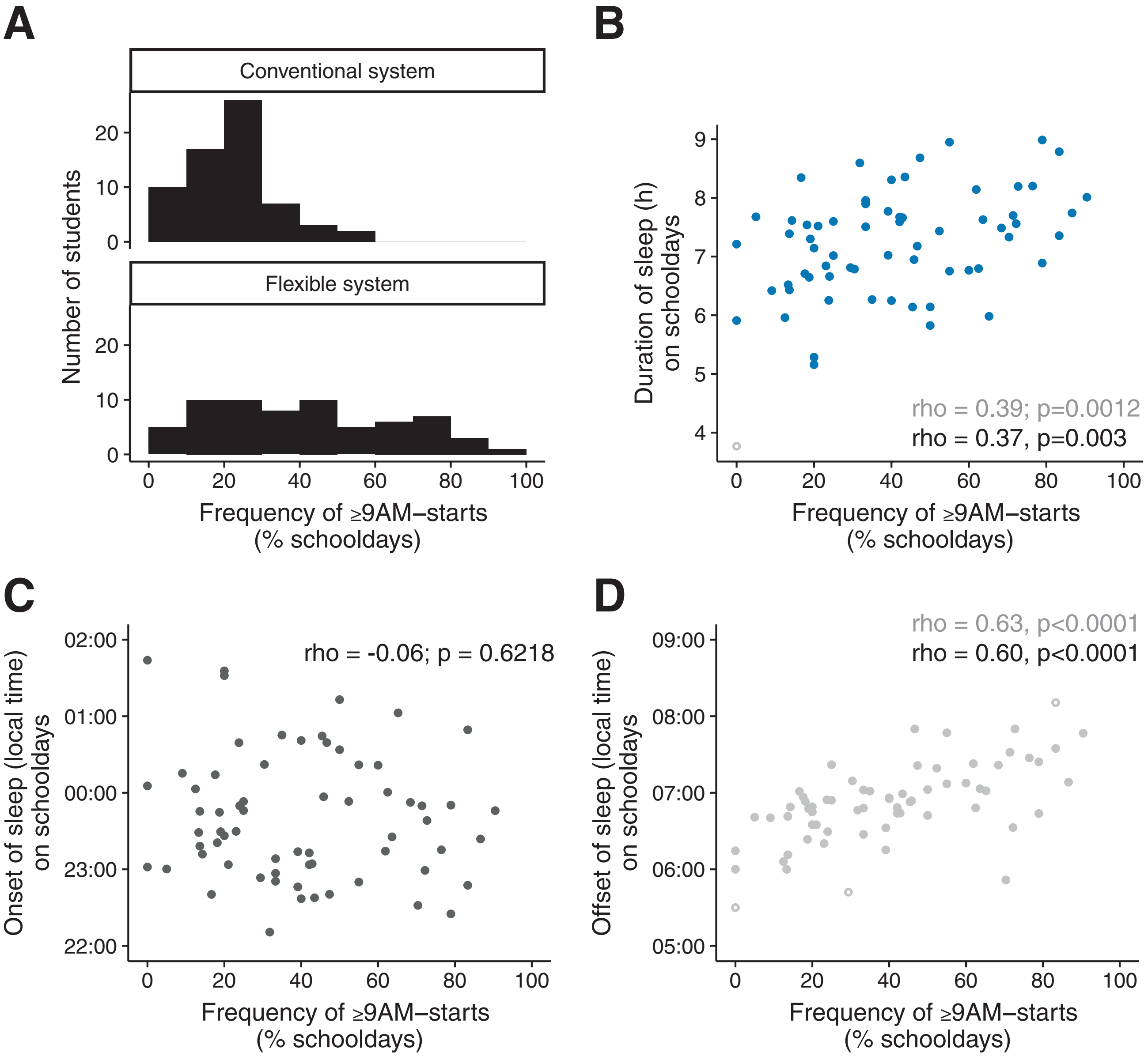

Figure 6. Extension of sleep in the flexible system across all schooldays in relation to the frequency of ≥9 am-starts. Sleep parameters are from sleep diaries of the full cohort ($n = 65$). The frequency of ≥9 am-starts is the proportion of schooldays that a student reported to have skipped the first period of i.e. attended school at ≥9 am. (A) Distributions of the frequency of ≥9 am-starts in the flexible system in comparison to that in the conventional system as retrospectively extracted from students' timetables (not exactly as in baseline due to teacher absences and exams). Average (B) sleep duration, (C) sleep onset, and (D) sleep offset times across all schooldays in the flexible system against frequency of ≥9 am-starts. Data are color coded as in Figure 3. Tukey outliers in y-axis values are marked by empty gray circles. Results of Spearman correlations are given for both data including outliers (gray) and excluding outliers (black). $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$ for Wilcoxon signed-rank test.

$p = 0.0012$) as well as their average sleep duration across the week ($\rho = 0.31$; $p = 0.0119$) (Figure 6B).

Importantly, this effect appears to be driven solely through later offset times because 9 am-use was highly correlated with wake-up time ($\rho = 0.63$, $p < 0.0001$) but not at all with sleep onset time ($\rho = 0.06$, $p = 0.6218$, Figure 6C and D). This suggests (although these are just associations) that going to school later more often does not delay overall sleep onsets, and hence the benefit is maintained. This is further supported by our finding that weekend sleep timing (onset, offset) and duration were also not associated with the frequency of ≥9 am-starts (weekend onset: $\rho = 0.06$, $p = 0.6584$; weekend offset: $\rho = 0.03$, $p = 0.7944$; weekend duration: $\rho = 0.06$, $p = 0.6368$).

Discussion

The debate about school start times is currently of very broad scientific and political interest given the widespread problem of teenage sleep deprivation. One of the first observations of a potential relationship between school start times and sleep was made in 1913 by Terman and Hocking [73]. They found that US students slept longer compared to German students; then schools started at 9 am in the US and at 8 am in Germany. This notwithstanding, school starts in the US have since become even earlier than those in Germany and those in Germany were maintained.

Evidence that this trend goes into the wrong direction has been accumulating over the last decades. Numerous studies have documented teenage sleep deprivation [e.g. 19–22, 74], linked it with short- and long-term performance and health deficits [e.g. 23, 40, 75, 76] and indicated that later school start times are likely an effective public countermeasure [e.g. 43, 44, 51, 77, 78]. However, most studies were performed on cross-sectional samples and are limited by nature in their design and thus evidence level [44, 45, 77] (with randomization and blinding virtually impossible). Hence, more studies with different designs and, particularly better sleep measures, are urgently needed.

We had the opportunity to monitor sleep intraindividually over many weeks in a group of German high-school students, whose school system was changed from a rigid one with mainly 8 am-starts to a flexible one with both 8 am- and ≥9 am-starts. Our results are in line with the majority of the other studies on school start times and support the need for a change in school schedules.

Sleep duration is longer on ≥9 am-days and the benefit is universal

In our study, virtually all participating students (97%) benefitted from later start times, sleeping longer on schooldays with a ≥9 am-start—on average students gained 1 h of sleep on those days. Importantly, not only was the overall benefit universal but also the magnitude of the benefit was similar across the important factors chronotype, gender, grade, and frequency of later starts. This may seem surprising at first but should actually be alarming: it exemplifies how severe and widespread teenage sleep deprivation may be, afflicting practically every single student leading to such ceiling effects. In our study sample, students rarely woke without their alarm clocks on schooldays, indicating that they rose before their sleep need was met and their internal day had started. Indeed, only 18% slept 8 h or more on their schooldays, the lower bound of the recommended 8–10 h for this population [79]—only 1 student slept on average over 9 h (based on sleep diary entries during baseline). These numbers are in line not only with the age and gender-matched adolescents across Germany used in our study (Supplementary Figure S1) but also with other studies in Germany [80] and around the world [e.g. 19–22, 74]—worrying statistics considering the acute and long-term health and performance detriments linked with teenage sleep deprivation [e.g. 23, 40, 75, 76].

On days with a ≥9 am-school start, these statistics looked much less bleak: 52% of students slept more than 8 h and 13% even more than 9 h, subjective sleep quality was improved and alarm-free waking was more likely (albeit still rare). However, the delay from an 8 am to a ≥9 am-school start was insufficient to separate the moderately sleep-deprived students from the heavily sleep-deprived students (and to bring out features linked with smaller or greater sleep gain). Similarly, the earlier chronotypes among the students were still quite late in their sleep timing compared to other age groups and thus benefitted as fully from the ~1-h delay as the later chronotypes. This suggests that the school start delay from 8 am to 9 am may be at the lower end of the required spectrum to counter teenage sleep deprivation.

Sleep onset does not delay

One of the greatest concerns regarding later school starts is that teenagers might be tempted to stay up even later in the evening either consciously or via delayed circadian rhythms from later exposure to advancing morning light. As a result, they would not gain more sleep but potentially further delay their circadian rhythms through prolonged exposure to delaying evening light. Supporting this line of thought, a recent modeling paper indicated that a delay in wake-up alone may not effectively increase sleep duration or reduce social jetlag long term (tested for 5 weeks in the model) unless controlling evening light exposure [81]. In our study, however, there was no evidence that sleep onset times differed between ≥9 am-days and 8 am-days. Even the students that went most often at ≥9 am did not show later sleep onsets than those that made less use of the 9 am-option. These findings tally generally with those from many other studies, which also did not observe systematic delays in adolescents' sleep timing after a delay in school starts [51]. However, also the opposite has been reported [51] and direct comparisons are hampered by the fact that many studies were based on cross-sectional data [45] and/or did not distinguish between bedtime (time of going to bed) and sleep onset (time of falling asleep). Importantly, the extended sleep and the stable sleep onset we observed in our study are based on a period of 6 weeks after the change into the new flexible system, suggesting that the sleep benefit might be maintained in the long term. However, longitudinal studies with follow-up assessments are needed to confirm this.

The flexible system: curse or cure?

There are many possible reasons for the absence of a delay in sleep onset. One reason may be the fact that school start choices influenced the opportunity for natural morning light exposure at the geographical location and season (winter/spring) of the study. During the flexible system, most students woke before sunrise on their 8 am-days and after sunrise on most of their ≥9 am-days. This longer window for natural daylight exposure before school—natural light is a stronger signal for the circadian system than artificial light—might have countered any circadian delay resulting from later timing of artificial light exposure at home in the morning. Alternatively, if considering psychological factors leading to stable sleep onsets, students reported to feel more alert and less tired and to sleep better in the flexible system. It is therefore possible that they consciously took advantage of longer sleep opportunities because they felt the benefits of getting more sleep.

However, also the flexibility of the system *per se* could be a reason for the stability in sleep onset. Even just knowing that one could wake up later if required might have improved students' attitude and anxiety around sleep, facilitating an earlier sleep onset and more restful sleep. Furthermore, variable wake up times may positively affect exposure to morning light from artificial sources (independent of sunrise times). *Permanent* later start times generally purport a delay in circadian timing by delaying overall light exposure in the morning for all days of the school week. In contrast, the occasional early start in the flexible system may help to prevent such a delay through ensuring occasional earlier light exposure. Therefore, one could speculate that providing flexibility may be instrumental in maximizing sleep benefits from later school starts—as long as increased sleep variability on schooldays can be offset by less sleep variability between schooldays and non-schooldays.

What argues against the positive impact of flexibility, however, is our finding that sleep duration was not significantly different between the conventional and the flexible system despite the clear sleep benefit when comparing ≥9 am-days to 8 am-days in the flexible system. The students in our cohort did not make great use of the 9 am -option but started school later on not even one full additional day per week, making it two days per week on average. With this low frequency of later starts, the net gain from the flexible system was negligible to nondetectable. This implies that students, given the choice, may not necessarily opt for what may be best for their sleep.

Why did students not opt for more later starts?

The low use of the 9 am-option greatly surprised us. It is not only at odds with the pervasive sleep deprivation in our sample but also with the results from our final survey where 64% indicated that an 8 am-school start was tough for them (always or most of the time) and 86% that a 9 am-start was actually easier (always or most of the time). Was the low frequency of 9 am-starts incorrectly reported or does it originate from a selection bias? Both seem unlikely: The low diary-reported 9 am-use from our participants tallied with the retrospective survey-reported 9 am-use—not only from the participants but also with that from additional 82 anonymous students that transitioned into the flexible system but did not take part in the study; students across the board did indeed not go later more often. Exploratory analyses did not reveal any stable predictors of 9 am-start frequency from baseline sleep, lifestyle or commuting factors (data not shown). However, given the many factors that can reasonably be assumed to influence the 9 am-use—of which many were not documented in our study (e.g. individual daily timetables, after-school appointments, carpooling, parents' attitude toward later school starts, exams, etc.)—our sample was likely insufficient for the complexity of the question.

Asked about the reasons for starting school at 8 am instead of 9 am in our survey, the most frequent answer (75% of students) was "to fulfil the school's quota of 10 self-study hours per week." If not enough free periods existed in a student's schedule, students had to stay longer in the afternoon. It is therefore likely that students opted for early mornings rather than late afternoons and thus made such little use of the 9 am-option. Time management training may help students to better organize their schedules in this regard, whereas the school may want to try to optimize their timetabling. Further frequent reasons for 8 am-starts were "easier logistics to get to school" (40%), an important factor in the implementation of changes in school start times, and "more time to learn (27%)," indicating that students got extra teacher-supervised study time when going to school at 8 am. Follow-up studies will hopefully shed light on this intriguing low use of later start times to guide better implementations of such a flexible system—which was after all liked by 98% of participating students.

Limitations

While selection bias is unlikely to explain the surprisingly low uptake of the 9 am-option as discussed above, it might still have had a systematic effect on some of our other results: of 253 eligible students, only 26% made up the final study cohort. Since the sleep characteristics of the study cohort closely match those of other German adolescents from (1) the MCTQ database sample and (2) other published data, the selection bias in this study is likely of a similar magnitude as in other studies.

All sleep durations in our study are based solely on nocturnal sleep of students. Occasional or regular naps were thus not considered in any of the analyses, which may have led to underestimation of total daily sleep duration in some students.

In addition, our assessment of alarm-driven waking might have underestimated the rate of non-natural waking since it did not cover students woken regularly by their parents or siblings. The detected decrease in alarm-driven waking on ≥9 am-days may thus not reflect only increased natural waking but also incorporate a switch from alarm-driven to parent-induced waking.

Finally, seasonal changes in photoperiod and associated changes in sleep timing and duration may have systematically influenced our findings, potentially explaining part of the null effect of the flexible system on sleep. With our study running from January to March at 50°N, our comparisons between baseline (Jan) and flexible system (Feb–Mar) were likely confounded by the gradual advance of dawn during spring linked with gradually earlier sleep offset times and shortening sleep durations [71, 72]. Therefore, seasonal changes in sleep may have offset potential positive (albeit small) effects of the flexible system on sleep rendering them undetectable. Vice versa, the one positive effect detected in the flexible system, the small reduction in social jetlag resulting from earlier timing of weekend sleep, might well be a false-positive finding caused by the seasonal trajectory toward earlier sleep. In contrast to these pre–post comparisons, comparisons of 8 am and ≥9 am-school starts within the flexible system were most likely independent of seasonal changes because 8 am and ≥9 am-days occurred interspersed and alternating throughout the flexible system within each individual.

Concluding remarks

Our study is one of the first evaluating the effects of later school start times on sleep and subjective performance parameters in Europe. A flexible system with both early and late start times could be a valid additional solution to the more common policy of delaying school start times outright—if students can be encouraged to use the late option frequently enough.

On days with a later start, students have the opportunity to sleep longer. This should reduce the accumulation of sleep debt during the week. The occasional 8 am-starts could be strategic in avoiding a delay in sleep onset by ensuring that students are exposed to light in the early morning on a weekly basis. In addition, especially important for practical applications, students prefer the flexible system and their subjective parameters are improved.

There are other examples of successful implementations of flexible school systems. In The Netherlands, there are schools where the main subjects are taught in the middle of the day (e.g. from 10 am to 2 pm), while students can choose whether to study minor, facultative subjects earlier in the morning or later in the afternoon. Such a system accommodates the wide distribution of chronotypes in the student population.

In conclusion, our results are in line with the accumulating scientific evidence supporting later school start times as a countermeasure against teenage sleep deprivation. We therefore urge more schools to delay their start times and to collaborate with scientists to increase our knowledge about the (long-term) effects of later starting times on sleep, subjective well-being, health, and performance.

Supplementary Material

Supplementary material is available at *SLEEP* online.

Acknowledgments

We thank Dr. LK Pilz for implementing the sleep diary online; A Khan Sullivan for school schedule digitization; the management of the Gymnasium Alsdorf, in particular W Bock and O Vollert, for their open support throughout the study; and the student participants for their willingness and effort to share their data. None of the authors have any private links with the Gymnasium Alsdorf, and the school was not involved in data analysis or interpretation nor the writing of the manuscript.

Conflict of interest statement. None declared.

Author Contributions

Conceptualisation: TR; Methodology: ECW, MV, LKP, CM, TR; Investigation: ECW, MV, CM, AMB; Data curation: CM, AMB, MV, AKS; Formal analysis: ECW; Validation: MV, DF, CM, AMB, GZ; Supervision: ECW, MV, TR; Visualization: ECW; Writing – original draft: ECW, AMB, GZ, CM; Writing – review and editing: ECW, MV, AMB, CM, GZ, DF, TR.

References

1. Dahl RE, *et al*. Importance of investing in adolescence from a developmental science perspective. *Nature*. 2018;**554**(7693):441–450.
2. Roenneberg T, *et al*. A marker for the end of adolescence. *Curr Biol*. 2004;**14**(24):R1038–R1039.
3. Crowley SJ, *et al*. A longitudinal assessment of sleep timing, circadian phase, and phase angle of entrainment across human adolescence. *PLoS One*. 2014;**9**(11):e112199.
4. Crowley SJ, *et al*. Sleep, circadian rhythms, and delayed phase in adolescence. *Sleep Med*. 2007;**8**(6):602–612.
5. Fischer D, *et al*. Chronotypes in the US – Influence of age and sex. *PLoS One*. 2017;**12**:1–17.
6. Crowley SJ, *et al*. An update on adolescent sleep: new evidence informing the perfect storm model. *J Adolesc*. 2018;**67**:55–65. doi:10.1016/j.adolescence.2018.06.001.
7. Carskadon MA, *et al*. An approach to studying circadian rhythms of adolescent humans. *J Biol Rhythms*. 1997;**12**(3):278–289.
8. Carskadon MA, *et al*. Regulation of adolescent sleep: implications for behavior. *Ann N Y Acad Sci*. 2004;**1021**:276–291.
9. Jenni OG, *et al*. Homeostatic sleep regulation in adolescents. *Sleep*. 2005;**28**(11):1446–1454.
10. Taylor DJ, *et al*. Sleep tendency during extended wakefulness: insights into adolescent sleep regulation and behavior. *J Sleep Res*. 2005;**14**(3):239–244.
11. Van den Bulck J. Television viewing, computer game playing, and Internet use and self-reported time to bed and time out of bed in secondary-school children. *Sleep*. 2004;**27**(1):101–104.
12. Munezawa T, *et al*. The association between use of mobile phones after lights out and sleep disturbances among Japanese adolescents: a nationwide cross-sectional survey. *Sleep*. 2011;**34**(8):1013–1020.
13. Cajochen C. Alerting effects of light. *Sleep Med Rev*. 2007;**11**(6):453–464.
14. Souman JL, *et al*. Acute alerting effects of light: a systematic literature review. *Behav Brain Res*. 2018;**337**:228–239.
15. Yang M, *et al*. The acute effects of intermittent light exposure in the evening on alertness and subsequent sleep architecture. *Int J Environ Res Public Health*. 2018;**15**(3):524. doi:10.3390/ijerph15030524.
16. Khalsa SB, *et al*. A phase response curve to single bright light pulses in human subjects. *J Physiol*. 2003;**549**(Pt 3):945–952.
17. Chang AM, *et al*. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proc Natl Acad Sci USA*. 2015;**112**(4):1232–1237.
18. Carissimi A, *et al*. School start time influences melatonin and cortisol levels in children and adolescents—a community-based study. *Chronobiol Int*. 2016;**33**(10):1400–1409.
19. Gibson ES, *et al*. “Sleepiness” is serious in adolescence: two surveys of 3235 Canadian students. *BMC Public Health*. 2006;**6**:116.
20. Matricciani L, *et al*. In search of lost sleep: secular trends in the sleep time of school-aged children and adolescents. *Sleep Med Rev*. 2012;**16**(3):203–211.
21. Keyes KM, *et al*. The great sleep recession: changes in sleep duration among US adolescents, 1991–2012. *Pediatrics*. 2015;**135**(3):460–468. doi:10.1542/peds.2014–2707.
22. Gradisar M, *et al*. Recent worldwide sleep patterns and problems during adolescence: a review and meta-analysis of age, region, and sleep. *Sleep Med*. 2011;**12**(2):110–118. doi:10.1016/j.sleep.2010.11.008.

23. Dewald JF, *et al.* The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: a meta-analytic review. *Sleep Med Rev.* 2010;**14**(3):179–189.
24. Hysing M, *et al.* Sleep and school attendance in adolescence: results from a large population-based study. *Scand J Public Health*. 2015;**43**(1):2–9.
25. Beebe DW, *et al.* Attention, learning, and arousal of experimentally sleep-restricted adolescents in a simulated classroom. *J Adolesc Heal*. 2010;**47**(5):523–525. doi:10.1016/j.jadohealth.2010.03.005.
26. Killgore WD, *et al.* Sleep deprivation reduces perceived emotional intelligence and constructive thinking skills. *Sleep Med*. 2008;**9**(5):517–526.
27. Vorona RD, *et al.* Dissimilar teen crash rates in two neighboring southeastern Virginia cities with different high school start times. *J Clin Sleep Med.* 2011;**7**(2):145–151.
28. Vaca F, *et al.* Drowsy driving. *Ann Emerg Med*. 2005;**45**:433–4. doi:10.1016/j.annemergmed.2005.01.015.
29. 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 (Lond)*. 2011;**35**(10):1308–1317.
30. Mullington JM, *et al.* Cardiovascular, inflammatory, and metabolic consequences of sleep deprivation. *Prog Cardiovasc Dis*. 2009. doi:10.1016/j.pcad.2008.10.003.
31. Raniti MB, *et al.* Sleep duration and sleep quality: associations with depressive symptoms across adolescence. *Behav Sleep Med*. 2017;**15**(3):198–215.
32. Short MA, *et al.* The impact of sleep on adolescent depressed mood, alertness and academic performance. *J Adolesc*. 2013;**36**(6):1025–1033.
33. Baum KT, *et al.* Sleep restriction worsens mood and emotion regulation in adolescents. *J Child Psychol Psychiatry.* 2014;**55**(2):180–190.
34. Tynjälä J, *et al.* Perceived tiredness among adolescents and its association with sleep habits and use of psychoactive substances. *J Sleep Res*. 1997;**6**(3):189–198.
35. Pasch KE, *et al.* Longitudinal bi-directional relationships between sleep and youth substance use. *J Youth Adolesc*. 2012;**41**(9):1184–1196.
36. Wittmann M, *et al.* Social jetlag: misalignment of biological and social time. *Chronobiol Int*. 2006;**23**(1-2):497–509.
37. Larcher S, *et al.* Impact of sleep behavior on glycemic control in type 1 diabetes: the role of social jetlag. *Eur J Endocrinol*. 2016;**175**(5):411–419.
38. Parsons MJ, *et al.* Social jetlag, obesity and metabolic disorder: investigation in a cohort study. *Int J Obes (Lond)*. 2015;**39**(5):842–848.
39. Roenneberg T, *et al.* Social jetlag and obesity. *Curr Biol*. 2012;**22**(10):939–943.
40. Owens J; Adolescent Sleep Working Group; Committee on Adolescence. Insufficient sleep in adolescents and young adults: an update on causes and consequences. *Pediatrics*. 2014;**134**(3):e921–e932.
41. Wheaton AG, *et al.* School start times, sleep, behavioral, health and academic outcomes: a review of literature. *J Sch Heal*. 2017;**86**:363–81. doi:10.1111/josh.12388.School.
42. Bowers JM, *et al.* Effects of school start time on students' sleep duration, daytime sleepiness, and attendance: a meta-analysis. *Sleep Health*. 2017;**3**(6):423–431.
43. Boergers J, *et al.* Later school start time is associated with improved sleep and daytime functioning in adolescents. *J Dev Behav Pediatr*. 2014;**35**(1):11–17.
44. Minges KE, *et al.* Delayed school start times and adolescent sleep: a systematic review of the experimental evidence. *Sleep Med Rev*. 2016;**28**:86–95.
45. Marx R, *et al.* Later school start times for supporting the education, health, and well-being of high school students. *Cochrane Database Syst Rev*. 2017;**7**:CD009467.
46. Lufi D, *et al.* Delaying school starting time by one hour: some effects on attention levels in adolescents. *J Clin Sleep Med*. 2011;**7**(2):137–143.
47. Nahmod NG, *et al.* Later high school start times associated with longer actigraphic sleep duration in adolescents. *Sleep*. 2019;**42**:1–10. doi:10.1093/sleep/zsy212.
48. Dunster GP, *et al.* Sleepmore in Seattle: later school start times are associated with more sleep and better performance in high school students. *Sci Adv*. 2018;**4**(12):eaau6200.
49. Lo JC, *et al.* Sustained benefits of delaying school start time on adolescent sleep and well-being. *Sleep*. 2018 ;**41**(6):zsy052. doi:10.1093/sleep/zsy052.
50. Carskadon MA, *et al.* Adolescent sleep patterns, circadian timing, and sleepiness at a transition to early school days. *Sleep*. 1998;**21**(8):871–881.
51. Wheaton AG, *et al.* School start times, sleep, behavioral, health, and academic outcomes: a review of the literature. *J Sch Health*. 2016;**86**(5):363–381.
52. Der Deutsche Schulpreis 2019. https://www.deutscher-schulpreis.de/preistraeger/gymnasium-der-stadt-alsdorf.
53. Parkhurst H. Education on the Dalton Plan. New York: E.P. Dutton & Company; 1922.
54. Dalton International n.d. https://daltoninternational.org/ (accessed June 11, 2019).
55. Roenneberg T. What is chronotype? *Sleep Biol Rhythms*. 2012;**10**:75–6. doi:10.1111/j.1479-8425.2012.00541.x.
56. Roenneberg T, *et al.* Epidemiology of the human circadian clock. *Sleep Med Rev*. 2007;**11**(6):429–438.
57. The WeP - the worldwide experimental platform n.d. https://www.thewep.org/documentations/mctq (accessed June 11, 2019).
58. Ghtobi N, *et al.* The μMCTQ - an ultra-short version of the Munich ChronoType Questionnaire. *J Biol Rhythms*. 2019. doi:10.1177/0748730419886986. [Epub ahead of print]
59. Roenneberg T, *et al.* Human activity and rest in situ. *Methods Enzymol*. 2015;**552**:257–83. doi:10.1016/bs.mie.2014.11.028.
60. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019.
61. Torchiano M. *R Package Effsize: Efficient Effect Size Computation*. Version 0.7.1 2017. https://cran.r-project.org/package=effsize.
62. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Heidelberg: Springer; 2016.
63. Kassambara A. *R package ggpubr: "ggplot2" Based Publication Ready Plots. Version 0.2* 2018. https://cran.r-project.org/package=ggpubr.
64. Harrell FE Jr. *R package Hmisc: Harrell Miscellaneous. Version 4.1-1* 2018. https://cran.r-project.org/package=Hmisc.
65. Bates D, Maechler M, Bolker B, Walker S. *R package lme4: Linear Mixed-Effects Models using "Eigen" and S4*. Version 1.1-18-1 2018. https://cran.r-project.org/package=lme4.
66. Kuznetsova A, Brockhoff PB, Christensen RHB. *R package lmerTest: Test in Linear Mixed Effects Models. Version 3.0-1* 2018. https://cran.r-project.org/package=lmerTest.
67. Pohlert T. *R package PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended*. Version 1.4.1 2019. https://cran.r-project.org/package=PMCMRplus

68. Neuwirth E. *R package RColorBrewer: ColorBrewer Palettes. Version 1.1-2* 2014. https://cran.r-project.org/package=RColorBrewer.
69. Wickham H. *R package reshape2: Flexibility Reshape Data: A Reboot of the Reshape Package. Version 1.4.3* 2017. https://cran.r-project.org/package=reshape2.
70. Rosenthal R. *Meta-analytic procedures for social research*. 2nd. CA:Sage; 1991.
71. Kantermann T, *et al*. The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Curr Biol*. 2007;**17**(22):1996–2000.
72. Hashizaki M, *et al*. A longitudinal large-scale objective sleep data analysis revealed a seasonal sleep variation in the Japanese population. *Chronobiol Int*. 2018;**35**(7):933–945.
73. Terman LM, Hocking A. The sleep of school childeen: its distribution according to age and its relation to physical and mental efficiency. *J Educ Psychol*. 1913;**4**:199–208.
74. Leger D, *et al*. Total sleep time severely drops during adolescence. *PLoS One*. 2012;**7**(10):e45204.
75. Roberts RE, *et al*. Sleepless in adolescence: prospective data on sleep deprivation, health and functioning. *J Adolesc*. 2009;**32**(5):1045–1057.
76. Dahl RE, *et al*. Pathways to adolescent health sleep regulation and behavior. *J Adolesc Health*. 2002;**31** (6 Suppl): 175–184.
77. Wahlstrom KL, *et al*. School start time effects on adolescent learning and academic performance, emotional health and behaviour. *Curr Opin Psychiatry*. 2017;**30**(6): 485–490.
78. Short MA, *et al*. A cross-cultural comparison of sleep duration between US and Australian adolescents: the effect of school start time, parent-set bedtimes, and extracurricular load. *Health Educ Behav*. 2013;**40**(3):323–330.
79. Paruthi S, Brooks LJ, D'Ambrosio C, Hall WA, Kotagal S, Lloyd RM, et al. Recommended amount of sleep for pediatric populations: a consensus statement of the American Academy of Sleep Medicine. *J Clin Sleep Med*. 2016;**12**:785–786. doi:10.5664/jcsm.5866.
80. Loessl B, *et al*. Are adolescents chronically sleep-deprived? An investigation of sleep habits of adolescents in the Southwest of Germany. *Child Care Health Dev*. 2008;**34**(5):549–556.
81. Skeldon AC, *et al*. The effects of self-selected light-dark cycles and social constraints on human sleep and circadian timing: a modeling approach. *Sci Rep*. 2017;**7**:45158.