Sleep: Educational Impact and Habit Formation*

Osea Giuntella
University of Pittsburgh
IZA and NBER

Silvia Saccardo
Carnegie Mellon University

Sally Sadoff
UC San Diego

September 30, 2023
(Click Here for the Latest Version)

Abstract

There is growing evidence on the importance of sleep for productivity, but little is known about the impact of interventions targeting sleep. In a field experiment among U.S. university students, we show that incentives for sleep increase both sleep and academic performance. Motivated by psychological theories of habit formation, our primary intervention couples personalized bedtime reminders to go to sleep with morning feedback and immediate rewards for sleeping at least seven hours on weeknights. The intervention increases average weeknight sleep by an estimated 19 minutes during a four-week treatment period with persistent effects of about 9 minutes per night during a one to five-week post-treatment period. Comparisons to secondary treatments show that immediate incentives have larger impacts on sleep than delayed incentives or reminders and feedback alone during the treatment period, but do not have statistically distinguishable impacts on longer-term sleep habits in the post-treatment period. We estimate that immediate incentives improve average semester course performance by 0.075 - 0.089 grade points, a 0.10 - 0.11 standard deviation increase. Our results demonstrate that incentives to sleep can be a cost-effective tool for improving educational outcomes.

JEL CODES: J10, I10, I23

Keywords: Sleep, Human Capital, Education, Habit Formation, Health Behaviors

*We are grateful to the seminar participants at the BEDI Conference, the Roybal Annual Retreat, University of Pittsburgh, the Center for Sleep and Circadian Rhythms Study, the Virtual Seminar on the Economics of Risky Behaviors (VERB), the CHIBE Behavioral Science and Health symposium, the University of Houston, the National University of Singapore, the Freie Universitat in Berlin, UCSD Spring School on Behavioral Economics, University of Chicago, RAND, University of Southern California, the UCLA Anderson School of Management, and the New York Federal Reserve. We are grateful to Mallory Avery, Daniel Banko-Ferran, Kelly Hyde, Ben Schenck, William Wang, Samuel Lindquist, and Gabrielle Toborg for their valuable research assistance. We are thankful to the Pittsburgh Experimental Economics Lab for recruiting our participants and the University of Pittsburgh Registrar Office for their help in providing access to students' academic records. We benefited from comments and discussion with Michele Belot, Alison Buttenheim, Gretchen Chapman, David Dickinson, David Huffman, George Loewenstein, Frank Schilbach, Heather Schofield, Peter Schwardmann, Kevin Volpp, Kareem Haggag. We received generous funding from J-PAL North America (Sadoff), the National Institute of Aging (Saccardo, Grant #P30AG034546) and the Pitt Healthy Lifestyle Institute Pilot and Feasibility project (Giuntella).

1 Introduction

There is growing attention to the role of sleep for economic outcomes (Hillman et al., 2006; Mullainathan, 2014). Sleep, an activity we all engage in every day and the single most common use of time, has been shown to influence cognition, physical health and mental well-being (Walker, 2017; Banks and Dinges, 2007). Statistics suggest that about 40% of Americans sleep less than the recommended minimum of seven hours per night and a similar proportion state that they would like to sleep more (Jones, 2013; Ballard, 2019). Roenneberg (2013) referred to sleep deprivation as the most prevalent high-risk behavior in modern societies. However, there is relatively little work on the causal impact of improved sleep habits on productivity and performance.

Using naturally occurring data, prior work argues that sunset and school start times affect labor market and educational outcomes via sleep (see Lindquist and Sadoff, 2023, for a review). These studies rely on quasi-random variation and either do not directly measure sleep or use self-reported survey measures, which can be biased (Lauderdale et al., 2008). In a separate line of research, sleep lab studies find that short-term severe sleep deprivation worsens attention, memory, cognition and mood, but are not able to estimate the effect of moderate longer-term changes in sleep in natural contexts (Killgore, 2010).

In recent work, Bessone et al. (2021) implemented a randomized field experiment in India that increased sleep through sleep aids and incentives. They find that the average increase of 27 minutes on nighttime sleep over a twenty day treatment period did not meaningfully improve cognition or productivity and could even decrease earnings because the increased time spent in bed decreased time spent working. The differential findings compared to those from naturally occurring data could be due to methodology: random assignment to an intervention directly aimed at increased sleep, carefully measured using sleep tracking devices. The differential findings compared to lab studies could be due to context: moderate increases in nighttime sleep occurring in an environment with low quality of sleep and among participants who are severely sleep deprived at baseline. Consistent with this, they find evidence that high quality sleep via office naps can increase productivity. The study cannot disentangle whether the differential effects of naps compared to nighttime sleep are due to the differences in the timing of sleep or the quality. It therefore remains an open question whether sustained exogenous increases in nighttime sleep can improve productivity and performance in the U.S., where average sleep is of high quality and closer to recommended guidelines.

Our study implements a randomized field experiment among U.S. college students to examine the impact that interventions targeting sleep have on sleep habits and academic performance. We ran the experiment over seven semester-long waves from Spring 2019 to

Spring 2022. The 1141 participants wore tracking devices (Fitbits) that measure sleep, heart rate and physical activity; downloaded a custom smartphone app linked to their Fitbit data that delivered our interventions; and, answered survey questions to capture information about their time use, cognitive performance and well-being. The study included a 1-4 week baseline period, followed by a four-week intervention period and a 1-5 week post-intervention period. In our treatment groups, we set a goal for participants to sleep at least seven hours by nine am on weeknights (Sunday - Thursday), following recommended guidelines (Panel et al., 2015).

Our intervention design builds on psychological theories of automatic cue-based habit formation, in which repeatedly coupling a consistent context cue with rewards for taking a desired action can create a positive association between the cue and the rewarded behavior (Wood and Rünger, 2016; Wood and Neal, 2007). Once this association is established, the cue may "automatically" trigger the desired action with little or no cognitive effort, even in the absence of a reward (Dickinson, 1985). Inspired by these theories, during the intervention period, participants in our primary treatment group received personalized reminders on weeknights, which served as cues to go to bed. On weekday mornings, they received feedback on whether they had met their sleep goal and, if successful, they received an immediate financial reward of $4.75. In the post-intervention period, we stopped the financial reward but continued to cue participants to sleep via bedtime reminders and provide morning feedback to reinforce the behavior. This allows us to investigate the persistence of behavior change without the presence of the reward. In secondary treatments, we tested variants that provided rewards with a delay rather than immediately; and, that turned off either the rewards or the cue and feedback. Our primary analysis compares a no intervention control group to the *Immediate Incentives* group, in which participants received bedtime cues, morning feedback and immediate incentives for each weeknight they met the sleep goal during the intervention period.[1]

At baseline, participants met the goal of sleeping at least seven hours on a little over 40 percent of nights. During the treatment period, the intervention increases the rate of sleeping at least seven hours on weeknights by an estimated 12 percentage points ($p < 0.001$), 28 percent higher than baseline. The treatment effects persist into the post-intervention period but are smaller: an estimated 5.6 percentage points ($p < 0.001$), a 13 percent increase

[1]The primary *Immediate Incentives* group pools two sub-treatments that received cues, feedback, and immediate incentives: one that continued to receive reminders and feedback in the post-treatment period and a secondary treatment group that did not receive reminders and feedback in the post-treatment period, which allows us to examine the importance of providing context cues for the persistence of behavior after the reward is removed. We do not find significant differences in the post-treatment effects of the two groups and pool them for our primary analysis.

compared to baseline. We estimate that average weeknight sleep increases by 19 minutes during the treatment period and 8 minutes during the post-treatment period ($p < 0.001$).[2]

We find that the intervention induced participants to adopt earlier bedtimes on average during the treatment period (and, directionally, earlier wake up times). However, this behavior does not persist in the post-treatment period: when rewards are removed, average bedtime largely returns to baseline levels with slightly later average wake-up times. While we see little persistence of treatment effects on average bedtime and wake up time across individuals, the intervention did lead to more regular bedtimes and wake up times within individuals, which are sustained at similar magnitudes in the post-intervention period. Taken together, these results suggest that the intervention led participants to establish more stable routines that may not have been reliant on the external cue to go to bed earlier.

To further explore mechanisms of habit formation, we compare our primary intervention to three secondary treatments: (1) *Delayed Incentives*, which is identical to Immediate Incentives but with the payout delayed to the end of the study (about a month after the treatment period); (2) *Delayed Incentives No Cue/feedback*, which is identical to Delayed Incentives except that participants do not receive cues or feedback; and, (3) *Cue/feedback*, which only provides reminders and feedback with no rewards. During the treatment period, we find that immediate incentives significantly outperform the secondary treatments. The effects of immediate incentives are 60 to 85 percent higher than delayed incentives and three to four times larger than cue/feedback alone. However, during the post-treatment period the effects of immediate incentives are not statistically distinguishable from any of the secondary treatments. Our results suggest that combining cues with immediate rewards has particularly large impacts while incentives are being offered but may not be particularly effective at enhancing the persistence of habits.

We then turn to the educational impact of incentives to sleep. Immediate incentives increase semester course performance by an estimated 0.075 - 0.089 grade points ($p = 0.044$ and $p = 0.035$, respectively).[3] We find evidence of similar sized treatment effects on grade point average in the semester following the intervention, but no effect two semesters after the intervention. We examine heterogeneity by course type and time of day and find that treatment effects are largest in STEM courses and classes that take place midday.

We benchmark our effects in comparison to prior work linking sleep to academic perfor-

[2]Our focus on weeknight sleep is in line with prior work that examines the impact of school and class start times, which occur on weekdays. We find no evidence of substitution between incentivized weeknight sleep and unincentivized sleep, including sleep that occurs during the day (i.e., naps), on weekends and during holidays.

[3]Effects are similar but with slightly smaller magnitudes if we pool all incentive treatments (immediate and delayed).

mance. Estimates from natural experiments in the U.S. suggest that a one hour later shift in sunrise or class start time increases sleep by an average of 6 - 35 minutes and has either no discernible impact on academic achievement or can increase grades and test scores by 0.06 - 0.16 standard deviations (SD) (Carrell et al., 2011; Heissel and Norris, 2018; Groen and Pabilonia, 2019). By comparison, our intervention increases weeknight sleep by an estimated 19 minutes during treatment and 9 minutes during post-treatment; and, grades by 0.10 - 0.11 SD.

We consider three primary channels through which sleep could affect academic performance: lifestyle, cognition, and well being. To that end, we examine the impact of our intervention on time use, performance in math and creativity tasks, and measures of physical and mental well-being. The intervention leads to declines in self-reported screen time, which includes internet browsing, TV/videos and games, and excludes screen time for studying. These changes are similar in magnitude to the increases in sleep and are concentrated around bedtime. We find that total study time does not change during the intervention period, but there is suggestive evidence of a reallocation of study time from evening hours to morning hours, with little change in other time use. We do not find treatment impacts on our measures of cognitive performance, math and creativity. We also do not find treatment effects on physical activity, or end of semester mental health, though we find evidence that treated participants are better able to cope with stress. Together, our results suggest that incentives to sleep lead to more regular sleep habits, which displace screen time, and shift study time to earlier hours of the day, which may contribute to the improvement in academic performance.

Our study is the first to show that an intervention targeting sleep can improve academic performance. These findings contribute to the growing literature on the economics of sleep. Seminal work on sleep finds a negative association in time use surveys between sleep and work hours, but is not able to identify causality (Biddle and Hamermesh, 1990; Basner et al., 2007). More recently, studies using naturally occurring data find that later sunset times are correlated with lower cognitive performance and earnings, and worse physical and mental health, arguing that the channel is via reduced sleep (Giuntella et al., 2017; Gibson and Shrader, 2018; Giuntella and Mazzonna, 2019; Jin and Ziebarth, 2020). Related work finds that earlier class times are correlated with less sleep and lower academic performance at both the K - 12 and post-secondary levels (Carrell et al., 2011; Heissel and Norris, 2018; Jagnani, 2021; Groen and Pabilonia, 2019). In addition, recent research among U.S. university students finds a strong positive correlation between academic performance and sleep, particularly in the first half of the term and for freshman (Creswell et al., 2023). However, Lusher et al. (2019) find no evidence of significant effects of sleep regularity on academic

outcomes in a large study at a Vietnamese university and little impact of delayed start times. None of these studies exogenously vary sleep or test policies to improve sleep habits.

Recent work using field experiments has tested the impact of interventions targeting sleep. In addition to Bessone et al. (2021) discussed above, two additional field experiments have tested the role of goal setting and commitment devices for improving sleep habits (Avery et al., 2022) and the impact of a cognitive behavioral intervention to address insomnia on negative affect, work satisfaction and self control (Barnes et al., 2017). There is also experimental evidence on how sleep deprivation affects decision making, ethical behavior and quasi-experimental evidence on voting (Dickinson and McElroy, 2017; Dickinson and Masclet, 2021; McKenna et al., 2007; Holbein et al., 2019).

Our findings also contribute to the large literature on improving academic performance, particularly among college students. Our intervention is highly cost effective compared to previously examined policies, including financial aid, mentoring and support services, and performance-based incentives. Our results suggest that it may be more cost effective to improve academic achievement via sleep rather than to incentivize performance directly. This finding is akin to recent work showing that incentives for exercise can improve educational achievement (Cappelen et al., 2017).

Finally, we contribute to the literature on habit formation by testing the impact of interventions that provide individuals with cues and reinforce behavior with rewards and repetition (Verplanken and Wood, 2006; Wood and Rünger, 2016). Related work in economics has examined evidence of cues generating automatic habits in the context of handwashing (Wellsjo, 2021) and promoted sustained behavior change using repeated feedback in the context of water conservation (Byrne et al., 2022). Additionally, prior research has shown that offering rewards for repeated engagement in desirable behaviors can promote sustained behavior change in the context of smoking, weight loss, exercise, and handwashing (Gneezy et al., 2011; Royer et al., 2015; Hussam et al., 2017; Beshears et al., 2021; Milkman et al., 2014). Building on these findings, Beshears et al. (2021) tested the impact of incentivizing exercise routines and, in contrast with the notion that stable routines are important for establishing habits, observed a negative impact of incentivizing exercise routines (versus offering flexible incentives to exercise) on habit formation. To our knowledge, none of this work has specifically examined the effect of combining consistent cues with immediate rewards on habit formation, which is made possible by our technology-based intervention. While existing studies have shown that immediate rewards can be more motivating than rewards received with even a small delay in the context of one-time behaviors (Levitt et al., 2016), prior work has not examined the sustained effect of immediate versus delayed rewards on repeated behaviors and post-intervention habit formation.

In the remainder of the paper, we describe the experimental design, data and analysis in Section 2. Section 3 presents the results and Section 4 concludes with a discussion of the cost-effectiveness of our intervention.

2 Experimental Design

We choose our experimental setting with three goals in mind. First, we aimed to accurately measure sleep in a natural environment outside the laboratory, in line with recent work (Bessone et al., 2021; Avery et al., 2022; Creswell et al., 2023). Second, we wanted to be able to causally assess the impact of sleep on sleep habits and academic performance, and capture additional measures of lifestyle habits, cognitive performance, and well-being. Third, we wanted to explore mechanisms that can effectively encourage positive sleep habits. To fulfill these goals, we focused on a population of college students at the University of Pittsburgh, measured sleep using wearable trackers, and designed a semester-long randomized controlled trial (RCT) aimed at improving sleep via financial incentives and other behavioral interventions, which we complemented with time diaries, and survey measures of cognitive performance and well-being. We linked the data collected in our experiment to administrative data on students' academic transcripts.

To gather objective measures of sleep in a natural setting, we had participants in our study wear personal wearable devices (Fitbits) that estimate sleep patterns based on movement and heart rate data. The use of such wearable trackers allowed us to move away from relying on sleep diary methods, which have been demonstrated to substantially overestimate sleep (Lauderdale et al., 2008; Bessone et al., 2021). Fitbits, which are among the most popular wearable trackers, are well-suited for monitoring sleep in natural settings due to their portability and unobtrusiveness, and are the most utilized wearables for biomedical research purposes (Wright et al., 2017). In our study, we used Fitbit Charge HR, Charge HR 2, Charge HR 3, Alta HR, and Inspire 2, which all capture both movement and heart rate. Recent studies have demonstrated the accuracy of these heart rate-enabled Fitbits compared to actigraphy, a commonly used method for outpatient sleep screening, suggesting their suitability for population-based sleep research (Haghayegh et al., 2019).

One source of concern with studies that rely on wearable trackers is that the devices require continued engagement via daily syncing. To ensure high sync levels for our experiment we developed a custom-made smartphone app that connected to the Fitbit API, which allowed us to monitor sync rates daily and notify participants with low sync rates to keep them engaged. The custom made app also allowed us to deliver our interventions to improve sleep habits via push notifications and the app itself, and to deliver feedback and immediate

rewards to participants upon meeting their sleep goals.

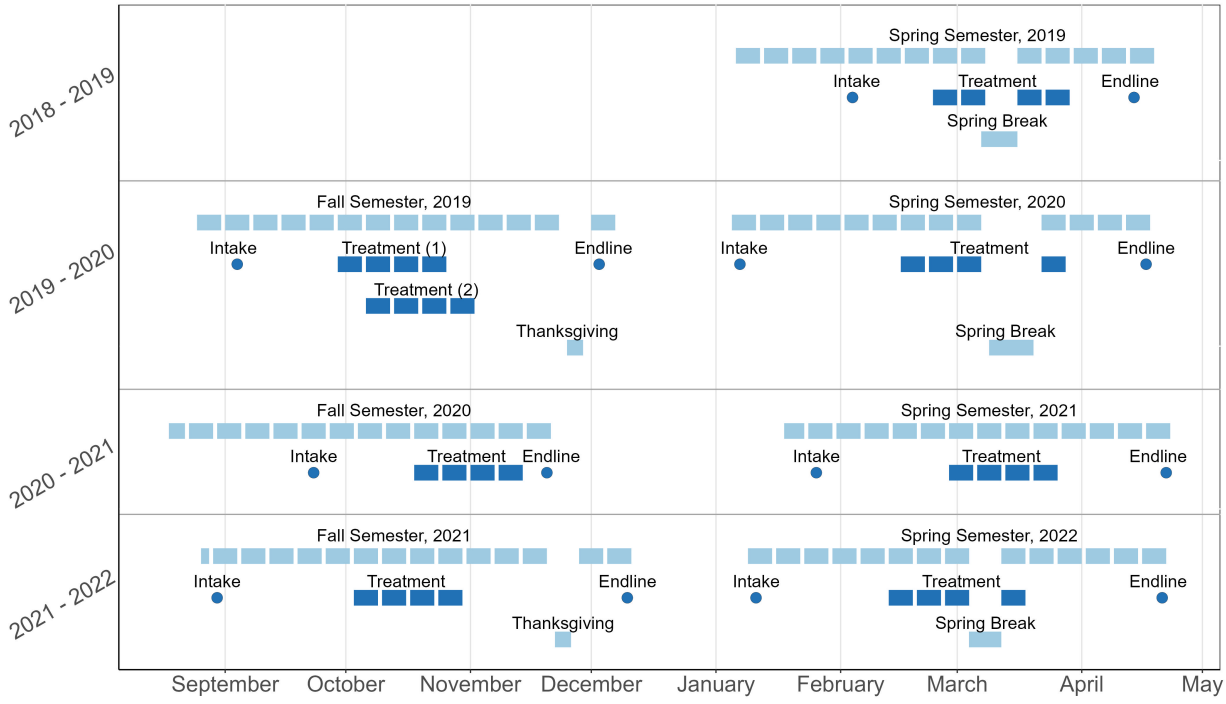
2.1 Sample, Recruitment and Timeline of the Experiment

The experiment took place at the University of Pittsburgh, was approved by the University of Pittsburgh Institutional Review Board and was pre-registered in the AEA registry (RCT ID AEARCTR-0003235). We recruited participants through the Pittsburgh Experimental Economics Laboratory (PEEL) and invited them to participate in a semester-long study on wellness for a guaranteed minimum payment of $30 and the opportunity to receive additional earnings based on luck as well as their choices during the experiment. To be eligible for our study, participants had to have a smartphone and be willing to wear and routinely synchronize a wearable device (Fitbit) during the semester. We began the experiment in Spring 2019 and enrolled participants every semester (Fall and Spring) until Spring 2022. We ran the experiment in seven consecutive waves, with modest-size cohorts to accommodate the number of participants we could recruit through the lab every semester as well as the number of Fitbits we had. Our final sample includes 1141 participants.

In each wave, the study lasted for approximately 10 weeks. We initiated participant recruitment in the first few weeks of the semester and enrolled participants in the experiment on a rolling basis. Upon enrollment, we measured baseline sleep for 1-4 weeks. At the end of the baseline period, we randomly assigned participants to either a control group or treatments designed to improve sleep habits, which lasted for 4 weeks (*Intervention Period*). Following our pre-registration, the control group and primary treatment group were conducted throughout the 7 waves of the experiment. In waves 1-3, we randomized participants to the control group, the primary treatment group, or one of several variations of the main treatment to understand mechanisms of habit formation. After the 4-week intervention period, we continued to follow participants for an additional 1-5 weeks until the end of the semester (*Post-Intervention Period*), at which point we asked them to return the Fitbit and fill out an endline survey. The study always ended during the last week of classes, before the final exams. In the different waves, the start of the recruitment period depended on lab availability. The timing of the treatment period and the length of the post-treatment period varied in each wave depending on when we were able to start recruiting, how quickly we enrolled participants, and the semester schedule. The timeline of the experiment for each of the seven waves is depicted in Figure 1.[4]

[4]In Spring 2020, the semester schedule was changed by the university closure prompted by the onset of the COVID-19 pandemic. Students enrolled in the study in Spring 2020 learned about the university moving to remote learning during spring break (mid-March 2020), and continued to stay enrolled in the study until the end of the semester. In the Appendix, we conduct sensitivity analyses that exclude the Spring 2020 wave.

Figure 1: Timeline of the Experiment



To enroll in the study, participants completed an initial session at the laboratory (Spring 2019 - Spring 2020) or over Zoom (Fall 2020 - Spring 2022), completed an intake survey, received the Fitbit and installed our custom smartphone app, which allowed us to track their Fitbit data and to interact with participants over the course of the study.[5] During the intake session, participants consented to wear and sync the Fitbit throughout the semester, answer weekly surveys, and grant us access to their academic records. They were informed about their right to withdraw from the study at any time with no penalty. To mitigate potential experimenter demand effects, we did not specifically disclose to participants that our interest was to study sleep behavior. Instead, we broadly explained that we were interested in wellness. Participants left the initial session with a one-page reminder outlining of what was expected from them during the study and agreed to return the Fitbit at the end of the semester. The intake survey administered to participants collected information on socio-demographic characteristics and baseline measures of physical and mental health.

Over the course of the study, participants in all treatments received reminders to sync their Fitbit via text message and the app (see Figure C.1). They also received weekly

[5]From Fall 2020 onwards, instead of filling out one unique survey during the intake session, the survey was split into an enrollment survey that participants filled out at enrollment while on Zoom, and a follow-up survey that was emailed to them a few days later. In Spring 2019-Spring 2020 and Fall 2021-Spring 2022, participants picked up the Fitbit from PEEL and received a $6 payment, In Fall 2020 and Spring 2021, participants received the Fitbit via mail.

surveys that elicited time use, cognitive performance and well-being. We describe the survey measures in more detail below (Section 2.3).

2.2 Treatments

After the baseline period, we randomized participants into treatment groups, which are displayed in Table 1.[6] Participants in the *Control* group (n=373, waves 1-7), received no intervention and continued to wear their wearable trackers and fill out surveys until the end of the semester. Participants in the treatment groups received interventions to improve sleep habits. In all treatments, we set the goal of sleeping at least 7 hours per night by 9am on weeknights (Sunday through Thursday).

Our main treatment, *Immediate Incentives* (N=467, waves 1-7), builds on the framework that habits are formed through repeated actions in response to specific context cues, with reinforcement playing a vital role in solidifying those habits. Cue-based theories of habit formation suggest that as individuals repeat a behavior in response to stable context cues, they develop automatic associations between the contextual cues and the behavior itself, and the presence of pleasurable rewards strengthens these associations (Hull, 1943; Wood and Neal, 2007; Verplanken, 2006). Sleep is an ideal domain for designing interventions that build on this framework, since it is a behavior that is repeated every day. To this end, the immediate incentives treatment combines cues to trigger action, immediate rewards and repetition.

To provide participants with a consistent cue, we sent them reminders to meet their target goal of sleeping 7 hours per night by 9 am every weeknight (Sunday-Thursday). These reminders had two major components. First, they included a personalized target bedtime calculated at the individual level based on each participants' baseline sleep patterns. We set a goal bedtime of about an hour earlier than their bedtime in the baseline period and sent them reminders to go to bed half an hour before their goal bedtime. Second, as the cue-based framework emphasizes the importance of a stable environment in triggering automatic behavior, we encouraged participants to engage in a specific bedtime behavior every weeknight before going to sleep. Participants selected their behavior from a menu of different options before the beginning of the intervention period. Examples included "Turn off your Phone", "Turn on bedtime music", Turn off your computer", "Turn on meditation

[6]We made two deviations from the pre-registration. Firstly, our original plan included an incentive treatment where participants would receive a $4.75 coupon for a breakfast treat at one of the University of Pittsburgh Einstein Coffee locations. However, due to unforeseen logistical issues, we had to suspend this treatment after the first few weeks of the first wave, and exclude it from the following waves. Secondly, the COVID-19 pandemic prevented us from meeting the pre-registered sample size for our main treatments, as we incurred additional costs for mailing Fitbits to participants.

app".[7] Appendix Figure C.1 displays the bedtime reminder.

Table 1: Treatments

	N	Waves	Treatment		Post-Treatment
			Reminders & Feedback	Rewards	Reminders & Feedback
Control	373	1-7	–	–	–
Immediate incentives	467	1-7	✓	Immediate	✓
Immediate incentives, post cue/feedback	356	1-7	✓	Immediate	✓
Immediate incentives, no post cue/feedback	111	5,7	✓	Immediate	–
Delayed incentives	103	1-3	✓	Delayed	✓
Delayed incentives, no cue/feedback	97	1-3	–	Delayed	–
Cue/feedback	101	1-3	✓	–	✓
Total	1,141				

Notes: The table reports the number of participants enrolled in each of the treatments; whether rewards were immediate or delayed, and whether they received reminders and feedback during and after the intervention. Immediate incentives pools Immediate incentives, post cue/feedback and Immediate incentives, no post cue/feedback.

Next, to link sleeping behavior with a pleasurable reward, we provided participants with immediate financial incentives upon meeting their sleep goal. Every weekday after 9 am, participants received feedback on whether they met their goal of sleeping at least 7 hours via the custom-made app through push-notifications and in the app interface. Participants who met their goal received feedback about having achieved the goal and earned a $4.75 reward, which they redeemed by clicking a button on the app (see Appendix Figure C.2). Participants in this treatment received a monetary reward of $4.75 through a Venmo transfer on the same day.[8] Participants who fell short of their sleep target were given feedback indicating that they had not achieved their goal and had missed out on receiving the reward. This feedback also included a negatively valenced emoji to convey the injunctive message that sleeping less than 7 hours was discouraged (see e.g., Schultz et al., 2007), and an encouragement to make another attempt to meet the sleep target. To encourage repetition, cues and rewards continued every weeknight and weekday of the 4-week intervention period. In waves in which the treatment period spanned spring break, we paused the intervention during spring break.

[7] On the Friday before the beginning of the intervention-period, participants received their intervention-related instructions. As part of these instructions, we asked participants to select a bedtime behavior to engage in before going to bed.

[8] For logistical reasons, the payment was received after 3pm each day, which introduced a delay between the performance of the behavior and the reward. However, the feedback about receiving a reward was provided as soon as participants synced the Fitbit after 9 am.

At the end of the intervention period, we discontinued the financial rewards.[9] In our main variation of the Immediate Incentives treatment, *Immediate Incentives – Post Cue/Feedback* (N=356, waves 1-7), we continued to send bedtime reminders (i.e., the cue) and morning feedback on whether they had achieved their sleep goal throughout the post-intervention period, which lasted through the end of classes. The feedback was identical to the post-intervention period, except that we removed mention of the financial reward, as displayed in Appendix Figure C.4). To examine the importance of maintaining the cue for habit persistence, we tested a variant of the immediate incentives treatment in which participants did not receive cue and feedback in the post-intervention period, *Immediate Incentives – No Post-Cue/Feedback* (N=111, waves 5 and 7). Our primary analysis pools the two variants of immediate incentives.

Secondary treatments. In the first three waves of the study, we implemented secondary treatments to examine the importance of cues and feedback, and the timing of financial incentives for habit formation.

In the *Delayed Incentives* treatment (N=103, waves 1-3), we provided participants with cues and rewards, as in the immediate incentives treatment. The only difference was that, although feedback about receiving the incentive was immediate, payment was not. Instead of receiving a $4.75 Venmo transfer each day they met their sleep goal, participants redeemed the incentive every day but only received one single transfer with the total payment at the end of the study period. Appendix Figure C.3 displays the feedback screens for the Delayed Incentives treatment. This treatment aimed to test whether providing repeated immediate rewards increases their effectiveness during treatment compared to delayed rewards, as has been shown with one-time rewards (Levitt et al., 2016); and whether reinforcing behavior with immediate rewards enhances persistence of behavior after the reward is removed relative to delayed rewards. In the post-intervention period, participants continued to receive the bedtime cue and morning feedback, as in the Immediate Incentives treatment.

In the *Delayed Incentives – No Cue/Feedback* treatment (N=97, waves 1-3), we removed the bedtime reminders (i.e., the cue) and the daily feedback about whether participants met their sleep goal but retained the financial incentive. At the start of the intervention period, we informed participants that they would receive $4.75 for every night they would meet the goal of sleeping at least 7 hours by 9 am, with payment to be received via a Venmo transfer at the end of the semester. This treatment is analogous to other work using financial incentives to create habits in the context of exercising (e.g., Charness and Gneezy, 2009; Royer et al., 2015) and aimed to test the importance of providing participants with cues and of clearly

[9]We notified participants at the end of the intervention period via text message.

linking the financial reward with the desired behavior (i.e., meeting the sleep goal). The participants did not receive reminders or feedback during the post-intervention period.

To test whether financial rewards are critical for establishing habits, we also conducted an additional treatment, *Cue/Feedback* (N=101, waves 1-3), where we removed the financial reward. Participants in this treatment received the same bedtime reminders as participants in the immediate incentives and delayed incentives treatments. They also received daily feedback via the app on whether they had achieved their sleep goal. Instead of providing participants with a reward, the feedback screen included a positively or negatively valenced emoji depending on whether participants had achieved their sleep goal – i.e., the same feedback that the Immediate and Delayed Incentives groups received in the post-intervention period (see Appendix Figure C.4). The participants in the Cue/Feedback treatment continued to receive the same reminders and feedback throughout the post-intervention period.

2.3 Outcome Measures

2.3.1 Sleep

Our primary pre-registered sleep outcome is the share of weeknights (Sunday—Thursday) participants sleep at least 7 hours. Our secondary measures of sleep include include sleep hours per night, sleeping between 7 and 9 hours on weeknights, bedtime, wake up time, sleep regularity, and sleep quality as measured by the Fitbit.

We define sleep regularity as the sleep variability across the week, as measured by the within-person standard deviation in the outcome. We measure sleep quality in terms of efficiency, Rapid Eye Movement (REM) sleep and deep sleep. Efficiency measures the percentage of time in bed that an individual is asleep. REM sleep is the stage of sleep in which individuals dream, which stimulates areas of the brain essential to learning. During REM sleep heart rate and blood pressure rise. Studies suggest that REM sleep plays a key role in memory consolidation, emotional processing, and brain development (Marks et al., 1995; Boyce et al., 2016). Deep sleep is the most restorative form of sleep. During deep sleep the heart rate and breathing rate are at their lowest and our body repairs tissue. Deep sleep is important for regulating glucose metabolism and has also been linked to cognitive function and memory (Zhang and Gruber, 2019; Leproult and Van Cauter, 2010).

We caution that while, as described above, there is growing evidence on the performance of recent Fitbit models in accurately measuring sleep duration (de Zambotti et al., 2018), the accuracy and reliability of these devices in capturing sleep stages needs further validation. In particular, sleep trackers have acceptable sensitivity but poor specificity when compared with sleep stages obtained using a polysomnography (PSG).

Sync rates throughout the study were relatively high. On average participants synced their devices for 88% of the days. In all of our analyses, for the nights with missing data, we replace missing data with an individual's baseline average following the approach of Bachireddy et al. (2019). We also conduct sensitivity analyses that do not replace missing data.

2.3.2 Educational outcomes

Our primary pre-registered educational outcome is individuals' term Grade Point Average (GPA), measured using administrative data. We obtained data on course grades and course credits from the University of Pittsburgh Registrar Office. We obtained data for each semester from Fall 2018 until the Spring of 2023 on the course grades of students enrolled in our experiment. We successfully matched 1120 students out of 1141 participants in our experiment (98.1% of the sample). The 21 unmatched students had no available grade for the term of the intervention, although they were successfully matched to their academic records in other terms. The data allow us to have information on the course grades for the term of the intervention and the following terms if the students were still enrolled in any class for a grade. We also obtained data from the Registrar on participants' GPA in High School, for which we could match 1032 students (90% of the sample). For the 71 of the remaining 109 students with no high-school GPA, we had information on baseline GPA in previous terms.

2.3.3 Additional outcomes

Time Use. We implemented a time use survey once a week, rotating the weekday on which the survey was administered. Our time use measure follows the structure of American Time Use Survey (Abraham and Flood, 2009). From a drop-down menu, participants indicated how they allocated their time on the previous day. For each 30 minute interval over the course of 24 hours, participants could choose from a list of activities that included sleeping, grooming (self), watching TV/videos, surfing the internet, playing games, working, studying, preparing meals or snacks, eating or drinking, cleaning, laundry, grocery shopping, attending religious service, hanging out with friends, paying bills, exercising, or commuting. They could also indicate that they did not know or could not remember how they spent their time, or could refuse to respond. In our primary analysis, we examine "screen" time, which pools time spent watching TV/videos, surfing the internet and playing games and excludes screen time spent studying; we categorize time spent hanging out with friends as "social" time.

Cognitive Performance. We measured cognitive performance through math and creativity

questions. We drew the math questions from the math section of the Graduate Record Examination Test (GRE) and we measured creativity using an adapted version of the task employed by Charness and Grieco (2019), where we provided participants with a list of 10 words and asked them to use some or all of the words to write an interesting sentence.. On alternate weeks, the weekly time use survey included either one multiple choice math question or the creativity task. Both tasks were incentivized (see instructions in Appendix E).

To measure creativity task, we recruited raters from the lab participants at the University of California San Diego and from Prolific (N=1,369). Raters received a random subset of the sentences produced by participants in the creativity task and rated them on a 1-5 scale. Each sentence was rated by a minimum of 2 raters; the median number of ratings per sentence is 8.

Physical Health. From the wearable trackers (Fitbits), we collected data on resting heart rate, daily steps, and active minutes. Resting heart rate measures heart beats per minute (BPM) at rest, i.e. when sitting, lying down or relaxing. Faster resting heart rates are associated with shorter life expectancy (Cooney et al., 2010; Dyer et al., 1980). Daily steps are the number of steps over the course of a 24 hour-period. Active minutes are measured as minutes in which a person is non-sedentary for a least 10 continuous minutes, where non-sedentary is defined as activity that raises heart rate enough to burn at least 3 metabolic equivalents (METs).[10]

Well-Being. To measure well-being, we collected measures of mood as well as measures of stress and ability to cope with stress (resilience). For mood, participants indicated, on a 10-point Likert scale, how happy they felt in that moment. For stress and resilience, participants indicated, using a 5-point Likert scale, 1) the extent to which participants faced stress in their life at the time of answering the survey and 2) the extent to which they felt able to deal with the stress they were facing. Every week we alternated between mood and the stress/resilience questions. These measures were collected via text message and, every week, participants were randomly assigned to receive the text message at different times of the day (11am, 4pm, 9pm). On top of these measures, we also collected measures of mental health in the intake survey (upon enrollment) and in the endline survey at the end of the semester. We measured depression using the Center for Epidemiologic Studies Depression scale (CES=D, Radloff, 1977), which is a 20-item validated instrument designed to assess the

[10]In practice, this measure sums the lightly active, fairly active and very active minutes collected by the Fitbit. Our pre-registered secondary measures of health also include Body Mass Index (BMI) and blood pressure, which we did not collect due to logistical constraints.

frequency of depressive symptoms on a scale from 0 (Rarely or None of the Time) to 3 (Most or all of the times). An overall depression score is calculated by summing answers to all 20 items. We also measured anxiety using the Generalized Anxiety Disorder scale (GAD-7, Williams, 2014), a 7-item scale designed to assess symptoms of generalized anxiety disorder. The instrument assesses the frequency of anxiety-related symptoms using a scale ranging from 0 (Not at all) to 3 (Nearly every day), with total scores ranging from 0-21.

2.4 Randomization and Baseline Characteristics

The randomization occurred at the end of the baseline period, the weekend before the start of the intervention-period. We employed a block randomized design, stratifying our participants by gender and the share of weeknights participants slept more than 7 hours (above vs below median).[11] In the initial waves of the study (Spring 2019 to Spring 2020) we randomized participants to one of five groups with equal probability (Control, Immediate Incentives, Delayed Incentives, Delayed Incentives no Cue/Feedback and Cue/Feedback).[12] For the remaining waves, following our pre-registration we randomly assigned participants to either the control group or the Immediate Incentives treatment, as we intended to scale up the most effective treatment. In Waves 5 and 7, we randomized participants in the Immediate Incentives treatment to either receive or not receive cue and feedback during the post-intervention period (Immediate Incentives - Post Cue/Feedback or Immediate Incentives - No Post Cue/Feedback).

Table 2 compares baseline characteristics in the Control group to the Immediate Incentives group. In the Control group (column 1), students slept at least 7 hours on only 44% of weeknights at baseline. On 29% of nights, students slept less than 6 hours. The average bedtime was ten minutes past 1am and the average wake up time was 5 minutes before 8am. The average high school GPA in our sample of 4.14 is representative of the overall University of Pittsburgh student population: the interquartile range of students offered admission at the University of Pittsburgh in 2022 had a weighted average GPA ranging from 3.91 to 4.42.[13] However, female students and Asian students are over-represented. The full-time Pitt population is made up of approximately 56% women, while women make up 73% of our sample. Asian students make up 11.4% of the student population, while they represent 27% of our sample. White students are slightly under-represented, 56% of our sample while making up 68% of the student population at Pitt, while the share of Hispanic and African-

[11] We did not balance the randomization on baseline GPA because GPA data was not available at the time of the randomization; we received GPA data once the study was completed.

[12] In Fall 2019, due to recruitment issues at PEEL, we recruited two groups of participants and had them start the intervention in a staggered way, as shown in Figure 1.

[13] Data available at: https://admissions.pitt.edu/first-year-student/class-profile/

Table 2: Treatment-Control Differences in Baseline Characteristics, Immediate Incentives

Variable	Control	Immediate incentives	Difference
Demographics			
Female	0.724	0.728	-0.005
	(0.448)	(0.445)	(0.031)
Age	19.463	19.343	-0.117
	(3.007)	(1.966)	(0.172)
White	0.556	0.567	0.012
	(0.498)	(0.496)	(0.035)
Asian	0.282	0.261	-0.012
	(0.451)	(0.440)	(0.031)
Black	0.087	0.069	-0.019
	(0.282)	(0.253)	(0.019)
Hispanic	0.038	0.054	0.012
	(0.191)	(0.225)	(0.015)
Other	0.038	0.049	0.007
	(0.191)	(0.217)	(0.014)
Highest parent ed less than college	0.260	0.285	0.022
	(0.439)	(0.452)	(0.031)
Highest parent ed college	0.287	0.289	0.002
	(0.453)	(0.454)	(0.032)
Highest parent ed more than college	0.453	0.426	-0.024
	(0.498)	(0.495)	(0.035)
Baseline sleep outcomes			
Sleep hours	6.625	6.658	0.012
	(0.958)	(0.903)	(0.064)
Sleep $\geq$ 7 hours	0.438	0.425	-0.020
	(0.276)	(0.275)	(0.019)
Sleep $\geq$ 6 hours	0.706	0.713	0.001
	(0.258)	(0.258)	(0.018)
Bedtime	25.265	25.211	-0.073
	(1.313)	(1.297)	(0.091)
Wake up time	7.956	7.935	-0.062
	(1.302)	(1.238)	(0.086)
Baseline academic characteristics			
Freshman	0.528	0.531	-0.010
	(0.500)	(0.500)	(0.032)
Sophomore	0.119	0.118	0.007
	(0.325)	(0.323)	(0.022)
Junior	0.228	0.212	-0.015
	(0.420)	(0.409)	(0.028)
Senior	0.125	0.139	0.019
	(0.331)	(0.347)	(0.024)
STEM major	0.583	0.570	-0.004
	(0.494)	(0.496)	(0.035)
Number of early classes	1.522	1.666	0.149
	(1.561)	(1.545)	(0.108)
High-School GPA	4.141	4.134	-0.006
	(0.438)	(0.433)	(0.032)
Baseline Term GPA	3.457	3.466	0.001
	(0.497)	(0.462)	(0.038)
Observations	373	467	840

Notes: The sample is restricted to individuals in the Immediate incentive treatment and individuals in the control group. All estimates in column 3 include wave fixed effects. Robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

American students is representative of the student population. Students in our sample are on average 19.4 years old, with a large share of freshmen (53%). Sophomore, junior, and senior students make up respectively 12%, 22% and 13% of our sample. 27% of the students in our sample report their parents' highest degree is less than a college education (either a high school degree and no college or a high school degree and some college), 29% report that at least one of their parents has a college degree and 45% report that at least one of their parents has a post-college degree.

In column 3, we estimate the treatment-control difference for each baseline characteristic from a regression that includes an indicator for the Immediate Incentives group and wave fixed effects. We do not find any statistically significant differences between average baseline characteristics in the Control group compared to the immediate incentives group. We also estimate Treatment-Control differences for each treatment group separately in Appendix Table A.2 and find statistically significant differences at the expected rate (e.g., about five percent of tests are significant at the $p < 0.05$ level).

2.5 Analysis

For outcome measures that are observed repeatedly throughout the study (e.g., nightly sleep), our primary regression analysis estimates treatment effects during the intervention period and the post-intervention period relative to the Control group. Formally, we estimate the following OLS model, unless otherwise noted:

$$Y_{it} = \beta_1 D_i * T_t + \beta_2 D_i * P_t + X_i + \rho_t + w_t + \mu_t + d_t + \epsilon_{it} \quad (1)$$

where Y_{it} is the outcome measure of interest; D_i is an indicator equal to one if an individual was assigned to the treatment group of interest; T_t is an indicator equal to one for any observation during the four-week intervention period; P_t is an indicator equal to one for any observation in the post-intervention period; X_i includes an individual's baseline value of the outcome variable, baseline sleep (percent of weeknights slept at least 7 hours),[14] indicators for the number of classes starting before 10am in a week (0-5), high-school GPA (in quartiles), previous term GPA (in quartiles) if high-school GPA is missing, and demographic controls for gender, age in years (dummies), race/ethnicity (Asian, Black, Hispanic, White, other), and indicators for parents' highest education (less than college (high school degree only or some college), college degree, or more than a college degree). For all individual characteristics, we included a missing indicator for whether the variable is missing. ρ_t, w_t, μ_t, d_t are a set of

[14]We exclude baseline sleep in regressions for sleep outcomes due to collinearity with the the baseline value of the outcome variable.

fixed effects for, respectively, wave of the experiment, week of the experiment, month of the year, and day of the week. Standard errors are clustered at the individual level.

For outcome measures that are observed only once during the study (e.g., course grades), we estimate the following OLS model, unless otherwise noted:

$$Y_{it} = \beta_1 D_i + X_i + \epsilon_{it} \tag{2}$$

where the variables are as described above. In regressions on course grades, the level of observation is the course weighted by the number of credits. Standard errors are clustered at the individual level.

Our main analysis compares the Control group to the primary treatment, the Immediate Incentives group. As pre-registered, we also present the analysis for the primary outcomes comparing the Control group to the pooled incentives treatments (see Appendix Tables A.4 and A.7). We report both unadjusted p-values and statistical significance adjusted for multiple hypothesis testing (MHT) within pre-registered families of secondary measures, using the method described in Anderson (2008).[15]

Our primary outcome measures are sleeping at least 7 hours and course grades during the intervention term. The families of secondary outcomes we consider are: effects of secondary treatments, secondary sleep outcomes, secondary educational outcomes, time use, cognitive performance, physical health, and mental well-being. The secondary treatments include tests of each secondary treatment compared to Control; and each secondary treatment compared to Immediate Incentives. The secondary sleep measures include: sleep hours, bedtime, wake up time, sleeping more than 6 hours per night, sleeping between 7-9 hours per night, sleep regularity, bedtime regularity, wake up time regularity, sleep efficiency, REM sleep and Deep sleep. The secondary educational outcomes include course pass rate. Time use includes all individual time use categories. Cognitive performance includes performance on the math and creativity tasks. Physical health includes heart rate, steps and physical activity. Mental well-being includes our measures of mood, stress, resilience to stress, anxiety (GAD - 7) and depression (CES - D). We compute the adjustment separately for treatment and post-treatment. In the Results section, we report unadjusted p-values and note which estimates are robust to adjustment.

[15]Note that adjusted p-values can be both larger or smaller than unadjusted p-values. This is because, as noted by Anderson (2008), sharpened FDR q-values can be less than unadjusted p-values when many hypotheses are rejected, because if you have many true rejections, then you can tolerate several false rejections too (this effectively just happens for p-values that are so large that you are not going to reject them regardless).

3 Results

We first examine treatment effects on sleep habits. We then turn to the impact of our intervention on educational outcomes. Finally, to explore potential mechanisms for our effects, we analyze time use, cognitive performance and health (physical activity and mental well-being).

3.1 Sleep

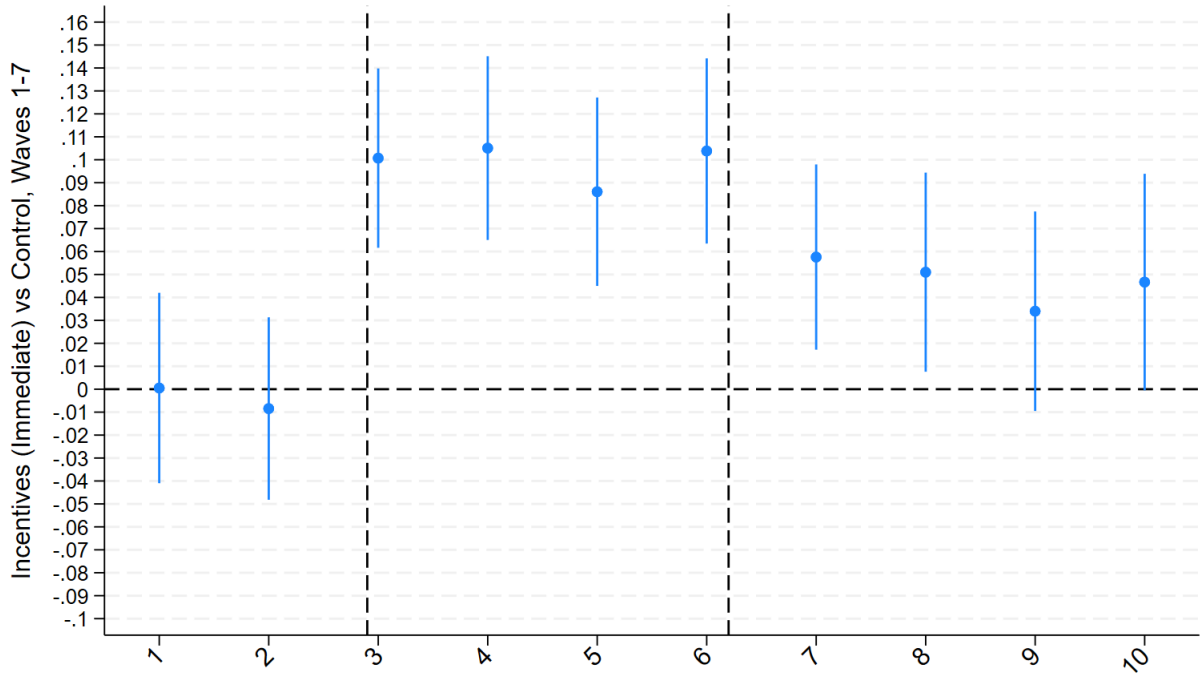
Baseline. Data from the baseline (pre-intervention) period reveals that a considerable portion of college students in our sample are sleep-deprived. In Panels A and B of Appendix Figure A.1, we present statistics on sleep duration and bedtime at baseline in the full sample, as measured by the Fitbit. Participants in our sample meet the recommendation of sleeping at least 7 hours approximately 43% of the nights (Panel et al., 2015). On average, at baseline participants sleep less than 6 hours approximately 28% of the nights, and less than five hours more than 10% of the nights. Panel B shows that about half of our participants have an average bedtime after 1 am, and about a quarter go to bed after 2 am on average. These data suggest that sleep deprivation is prevalent in our sample and is in line with a recent report by the National Institutes of Health indicating that more than 70% of college students sleep less than eight hours a day (Hershner and Chervin, 2014).

In the intake survey, we asked participants how many hours they slept on a typical weeknight and how many hours of sleep they considered optimal (see Appendix Figure A.1 Panel C). Consistent with previous studies, participants tend to report longer sleep than what is measured using Fitbit data (Avery et al., 2022; Lauderdale et al., 2008). 67% of the participants report sleeping at least 7 hours on a typical weeknight of the term (vs. 44% measured by the Fitbit). Yet, even when comparing to their self-reported sleep, 82% of the participants state a longer optimal sleep duration during the week. On average, participants' stated optimal sleep time is an hour longer than what they report as their typical sleep duration. 98% of the students in our sample report that their optimal sleep on a weeknight would be at least 7 hours, and 80% report an optimal sleep time during the week longer than 8 hours.

Intervention and Post-Intervention Period. We first estimate treatment effects on the primary incentivized measure of sleep: sleeping at least 7 hours by 9 am on weeknights (Sunday - Thursday). This outcome variable only includes nighttime sleep, and excludes weekends, holiday and naps (defined as episodes of sleep that start between 7am-8pm). Figure 2 plots the estimated difference in the rate of sleeping at least 7 hours between the treatment (Immediate Incentives) and Control groups, by week. The estimates are from

regressions by week in which individual-nights are the level of observation and we include an indicator for the treatment group with no additional covariates (the control group is the omitted group). Standard errors are adjusted for clustering at the individual level (the bars in the figure indicate 95% confidence intervals).[16]

Figure 2: Immediate Incentives and Sleep ≥ 7hrs (Weeknights), Excluding Naps



Notes - The sample is restricted to weeknights (Sunday-Thursday nights). On the horizontal axis we report week of the study: baseline (weeks 1-2), treatment (weeks 3-6), post-treatment (weeks 7-10). The coefficient reports the difference in the likelihood of sleeping at least 7hrs between individuals in the Immediate Incentives treatment and those in Control by week. Standard errors are clustered at the individual level. Bars indicate 95% Confidence Intervals.

In the first two columns of Table 3 Panel A, we present regression estimates of the treatment and post-treatment impacts of immediate incentives on sleeping at least 7 hours on weeknights and weeknight sleep hours, following the specification described in equation 1. At baseline, participants meet the goal of sleeping at least seven hours approximately 43% of the nights.[17] During the intervention period, immediate incentives increased the rate of

[16] We present the analogous figures for sleep hours in Appendix Figure A.2 and distributions of sleep hours in Appendix Figure A.3.

[17] The baseline average reported in Table 3 is slightly different from that reported in Table 2. This is because in Table 2 we calculate the baseline average at the individual level and in Table 3 we calculate it at the night level and not all participants have the same number of nights in the baseline period due to rolling

sleeping at least seven hours by an estimated 11.8 percentage points, a 27.5% increase. The treatment effects persist into the post-intervention period but are about 53 percent smaller: an estimated 5.6 percentage points, 13.1% higher than baseline.[18]

We estimate that total sleep hours increase an estimated 19 minutes on average during the intervention period and an estimated 8 minutes during the post-intervention period. All estimates are significant at the $p < 0.001$ level and are robust to adjusting for multiple hypothesis testing, except for the post-treatment effect on sleep hours (adjusted $p = 0.232$).

As shown in Appendix Table A.3, we do not find any evidence of substitution between incentivized weeknight sleep and unincentivized sleep during the day, on weekends or during holidays (spring break for the treatment period and Thanksgiving for the post treatment period). If anything, we find small positive spillovers, with some evidence of an increase in the likelihood of sleeping at least 7 hours during weekends in the post-intervention period.[19]

In Appendix Table A.3 we also examine additional sleep outcomes. Similar to our main results, we find that our intervention significantly increases the share of nights participants sleep at least six hours and the share of nights they sleep 7-9 hours, with persistent but smaller impacts in the post-treatment period.[20] On our measures of sleep quality, we find small positive increases in minutes of REM sleep, small, marginally significant impacts on sleep efficiency and no significant impact on minutes of deep sleep. We note that in our sample baseline efficiency is high: participants are asleep an estimated 94% of the time they are in bed. By comparison, Bessone et al. (2021) estimate efficiency of 70% among their experimental participants in India.

To examine heterogeneity, we estimate treatment effects separately across the four quartiles of the distribution of share of weeknights with more than 7 hours of sleep at baseline (Appendix Table A.5). During the intervention period, we estimate large and significant treatment effects throughout the distribution. We find some evidence of heterogeneity in the post-treatment period, with larger effects among those with lower levels of sleep at baseline. These results suggest that our intervention has the most persistent impact on those who are

enrollment.

[18] As discussed in Section 2, our primary analysis replaces missing data with an individual's baseline mean. Appendix Table A.4 reports the effects of our intervention when we do not replace missing data with baseline means and the effects are similar, if not larger.

[19] Our estimate of the impact of our intervention on any sleep episode including naps, holidays and weekends shows that the intervention increased the share of nights with at least 7 hours of sleep by 7.2 percentage points in treatment and 4.4 percentage points in post-treatment, and increased total sleep hours by an estimated 13 minutes in the treatment period and 6 minutes in the post-treatment period ($p < 0.01$ for all estimates).

[20] Sleeping less than six hours is a common metric of sleep deprivation (Hafner et al., 2017). The recommendation of sleeping seven to nine hours draws on studies that link excessive sleep duration to detrimental effects on health (Hirshkowitz et al., 2015; Jike et al., 2018).

Table 3: Immediate Incentives and Sleep

	Panel A: Incentives and Sleep			
	sleep $\geq$ 7 hrs	sleep hours	bedtime	wake up time
Treatment	0.1182***	0.3206***	-0.3152***	-0.0477
	(0.013)	(0.035)	(0.037)	(0.034)
Post-Treatment	0.0556***	0.1436***	-0.0285	0.0704*
	(0.015)	(0.037)	(0.037)	(0.038)
Observations	46,989	46,989	46,989	46,989
R-squared	0.256	0.267	0.592	0.519
Mean of Dep. Var.	0.429	6.647	25.22	7.939
Std. dev.	0.495	1.279	1.457	1.438
Number of individuals	840	840	840	840
	Panel B: Incentives and Sleep Regularity (Weekly Analysis)			
		Regularity (within-individual weekly s.d.)		
		sleep hours	bedtime	wake up time
Treatment		-0.1033***	-0.0441*	-0.0654**
		(0.032)	(0.023)	(0.031)
Post-Treatment		-0.0474	-0.0536**	-0.0911**
		(0.035)	(0.025)	(0.036)
Observations		8,631	8,631	8,631
R-squared		0.035	0.033	0.038
Mean of Dep. Var.		1.171	0.896	0.937
Std. dev.		0.614	0.435	0.534
Number of individuals		840	840	840

Notes: The sample is restricted to individuals in the Immediate Incentive treatment and individuals in the Control group. All estimates include day of the week, week of the experiment, wave, and month fixed effects, baseline value of the outcome variable, and demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), indicators for the number of classes starting before 10am, indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all characteristics, we included a missing indicator for whether the variable is missing. In Panel A, observations are the dependent variable at the daily level. In Panel B, observations are the standard deviation of the dependent variable at the weekly level. Standard errors are clustered at the individual level. Mean of Dep. Var. is the mean of the dependent variable at baseline. Std. dev. is the standard deviation of the dependent variable at baseline.

most sleep deprived.

3.1.1 Drivers of Short and Long-term Sleep Habits

Bedtime and wake up time. In the last two columns of Table 3 Panel A, we estimate treatment effects on bedtime and wake up time. During the intervention period, treated participants go to bed an estimated 19 minutes earlier than participants in the control group ($p < .001$) with directionally earlier wake up times (column 4, $p = .163$). This pattern does not persist in the post-treatment period when the incentives ended but the bedtime reminders continued. Instead, average bedtime largely reverts to baseline levels and treated participants wake up slightly later ($p = .067$). As shown by Appendix Figure A.2, both bedtime and wake up time get progressively later over the course of the intervention period and stabilize during the post-intervention period. These results suggest that combining bedtime reminders with incentives induces participants to go to bed earlier, but does not establish a sustained habit linked to the bedtime cue that persists once incentives are removed.

Sleep Regularity. Panel B of Table 3 estimates treatment and post-treatment effects on sleep regularity. To do so, we examine the within-individual standard deviation of total sleep hours, bedtime and wake up time at the week level (the level of observation is individual-week).[21] We find significant decreases in sleep variability across the week, equivalent to about 10 percent of baseline, or 0.1 standard deviations. The magnitude of the effects during the intervention and post-intervention periods are of similar size and are significant at the 5% level after adjusting for multiple hypothesis testing. These findings suggest that, while treated participants do not on average sustain earlier bedtimes after the intervention ends, they do develop more regular bedtime and wake up time habits that persist into the post-intervention period, potentially by shifting to bedtimes and wake up times they are able to maintain. Such regularity may be important for cognition and performance. Prior work suggests that irregular sleep and light exposure patterns among college students are associated with delayed circadian rhythms and lower academic performance (Phillips et al., 2017; Trockel et al., 2000; Smarr, 2015).

Secondary Treatments. As discussed in Section 2 and summarized in Table 1, our three pre-registered secondary treatments vary elements of our primary Immediate incentives treatment in order to investigate the importance of cues and immediate rewards: (1) Delayed incentives, which is identical to Immediate Incentives except that the rewards are distributed at

[21] The regressions follow the specification of equation 1, except we exclude day of the week fixed effects given the analysis is at the weekly level.

the end of the study about a month after treatment; (2) Delayed incentives No Cue/Feedback, which is identical to delayed incentives except that participants do not receive a cue or feedback; and, (3) Cue/Feedback which only provides cues (bedtime reminders) and feedback with no rewards.

Table 4 estimates the treatment and post-treatment effects of all treatments separately on our main sleep outcomes, restricting the analysis to waves 1-3 of the experiment, when the secondary treatments were conducted. As shown in columns 1-2, the effects of immediate incentives during treatment are about 53% to 86% percent higher than the effect of delayed incentives (with or without reminder and feedback); and about three to four times higher than the effects of reminders and feedback alone. The differences between the estimated impact of immediate incentives and each of the secondary treatments are all significant at the $p < 0.05$ level after adjusting for multiple hypothesis testing.

During the post-intervention period, the estimated effects of immediate incentives are statistically indistinguishable from any of the secondary treatment groups. The estimated magnitude of the post-treatment effects of immediate incentives are substantially larger than those of delayed incentives or cue/feedback alone. Interestingly, the delayed Incentives no cue/feedback treatment, which similarly to prior work on the impact of incentive on sustained behavior change (e.g., Charness and Gneezy, 2009) promised end-of-semester incentives for meeting the goal of sleeping 7 hours of sleep but did not attempt to increase the persistence of habits through daily bedtime reminders and immediate rewards, has similar impact on post-treatment sleep as the Immediate Incentive treatment.

In Appendix Table A.4, columns 6 and 7, we estimate the post-treatment effects of the immediate incentives intervention separately for the subgroup of participants who continued to receive personalized bedtime reminders and feedback in the post-treatment period (immediate Incentive with Cue/Feedback) and for participants who stopped receiving them at the end of the intervention-period (immediate incentive no cue/feedback). Our estimates reveal no significant differences between these two subgroups, suggesting that receiving bedtime cues after incentives stopped did not further help sustain the routines developed during the intervention period.

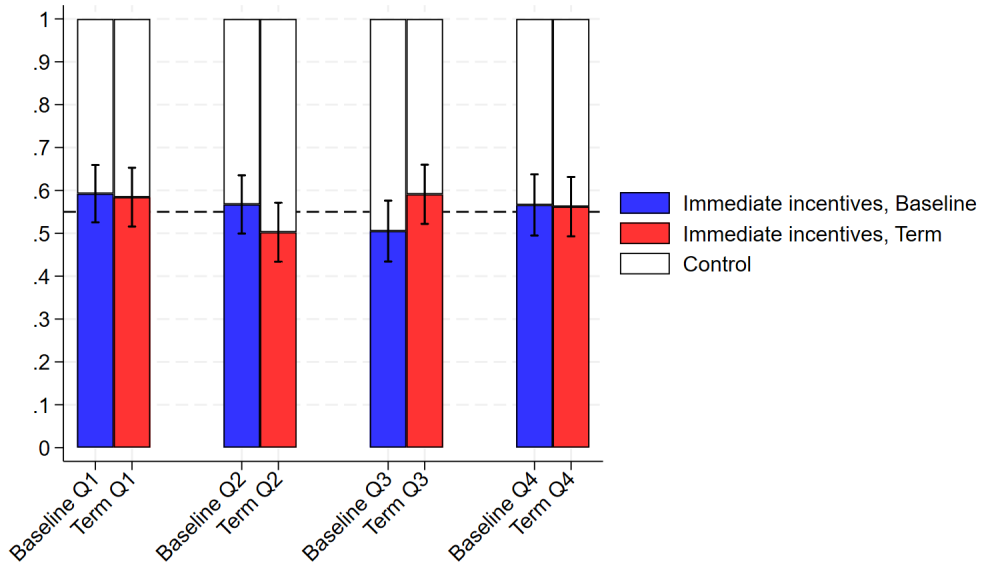
Collectively, the results presented in this section do not provide strong evidence that individuals built automatic habits as a result of our external cue, as evidenced by the fact that in the post-intervention period average bedtime reverted to baseline and that we find little impact of coupling bedtime cues with immediate rewards as compared to variations of the intervention that do not. Nonetheless, the observed increase in sleep regularity, with reduced variability in sleep hours, bedtime and wakeup time that persists in post-treatment, suggests that the intervention helped participants develop more consistent sleep routines

Table 4: Secondary Treatments

	sleep ≥ 7	sleep hours
Treatment		
Immediate incentives	0.1447***	0.3299***
	(0.021)	(0.054)
Delayed incentives	0.0946***	0.1777***
	(0.023)	(0.059)
Delayed incentives, no feedback	0.0893***	0.1778***
	(0.022)	(0.059)
Cue/Feedback only	0.0358*	0.1160**
	(0.020)	(0.052)
Post-Treatment		
Post: Immediate incentives	0.0503**	0.2120***
	(0.025)	(0.067)
Post: Delayed incentives	0.0139	0.0404
	(0.027)	(0.073)
Post: Delayed incentives, no feedback	0.0553**	0.0913
	(0.028)	(0.076)
Post: Cue/Feedback only	0.0134	0.0803
	(0.027)	(0.074)
Observations	34,954	34,954
R-squared	0.250	0.249
Mean of Dep. Var.	0.434	6.695
Std. dev.	0.496	1.538
Number of individuals	589	589

Notes: The sample is restricted to waves 1-3. All estimates include day of the week, week of the experiment, wave, and month fixed effects, baseline value of the outcome variable, indicators for the number of classes starting before 10am, and demographic controls for gender, age (dummies), race/ethnicity (Asian, Black, Hispanic, White, other), indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all characteristics, we included a missing indicator for whether the variable is missing. For all demographic characteristics, we included a missing indicator for whether the variable is missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. Std. dev. is the standard deviation of the dependent variable at baseline.group at baseline.

Figure 3



Notes - The figure reports the share of individuals in each quartile of the GPA distribution for both baseline GPA (the high-school GPA) and the term GPA during the intervention for the immediate incentive treatment. Bars indicate 95% Confidence Intervals.

independent of the external cue we provided.

3.2 Educational Outcomes

Next, we investigate the impact of our primary treatment on educational outcomes.

Figure 3 displays the share of individuals in the Immediate Incentives treatment and in the control group who are in each quartile of the GPA distribution at baseline and at the end of the intervention term.[22] The figure highlights that, as compared to baseline GPA, the share of treated participants in the bottom two quartiles declines and the share in top two quartiles increases, with the biggest shift in the middle two quartiles. Table 5 presents regression estimates of the impact of the intervention on semester GPA and secondary educational outcomes.[23] The regressions follow the specification for equation 2. In columns 1- 2 we estimate treatment effects on our primary outcome, course grade, in the term the intervention took place. Column 1 includes all course types (lectures, seminars, labs, independent studies

[22] We use high school GPA for baseline and only include participants with high school GPA in the figure.

[23] As discussed in Section 2.5, we are missing GPA for 4.6 percent of our participants. We examine differential attrition on the GPA measure in Appendix Table A.1 and find no evidence for differential attrition on term GPA.

Table 5: GPA, Immediate Incentives

	(1)	(2)	(3)	(4)	(5)	(6)
	Term of intervention		Term +1		Term +2	
	All classes	Lectures	All classes	Lectures	All classes	Lectures
Immediate incentives	0.075**	0.089**	0.070*	0.093**	0.006	0.006
	(0.037)	(0.042)	(0.038)	(0.042)	(0.042)	(0.046)
Observations	4,256	3,382	4,053	3,272	3,805	3,050
R-squared	0.129	0.137	0.095	0.107	0.085	0.090
Mean of Dep. Var.	3.499	3.434	3.552	3.493	3.545	3.501
Std. dev.	0.766	0.807	0.757	0.796	0.776	0.808
Number of individuals	825	819	778	776	720	711

Notes: All estimates include demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), baseline sleep, indicators for the number of classes starting before 10am, indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline.

and other classes) whereas Column 2 restricts the analysis to lectures (which account for approximately 80% of course types). Columns 3-4 report the same analysis for the term following the intervention. In columns 5 and 6, we examine the persistence of the effects two terms after the intervention. [24]

As shown in columns 1 and 2, we estimate that immediate incentive improved course performance by 0.075-0.089 grade points on average in the semester of the intervention ($p = 0.044$ for column 1 and $p = 0.035$ for column 2). We estimate a treatment effect of similar magnitude on course performance in the semester following the intervention, providing suggestive evidence of persistent effects (columns 3 ($p = 0.066$) and 4, $p = 0.027$). However, we do not find treatment effects in the semester two terms after the intervention (columns 5 and 6).

The estimated GPA impacts of 0.075 - 0.089 grade points during the intervention semester are equivalent to a 0.10 - 0.11 standard deviation (SD) increase in grades. In Appendix Table A.7 we report various sensitivity checks on course grade. Interestingly, when we exclude the Spring 2020 semester (wave 3), which experienced the onset of the COVID-19 pandemic leading to the abrupt closure of the university in the middle of the semester and

[24]Following our pre-registration, in the Appendix we pool all incentive treatments and repeat the main analyses; the results are similar with slightly smaller average impacts, as shown in Table A.4 for sleep and Table A.7 for educational outcomes.

disruptions in sleep and other lifestyle habits (Giuntella et al., 2021), our estimated treatment impact on term GPA is slightly higher, an estimated 0.091 grade points or 0.11 SD and 0.106 grade points or 0.13 SD when restricting to lectures.

Table 6: Immediate Incentives and GPA, Heterogeneity by Schedule and Class Type

	(1) Course grade	(2) Class Start: before 10am	(3) Class Start: 10am-2pm	(4) Class Start: after 2pm	(5) Class Type: non-STEM	(6) Class Type: STEM
			Panel A: All classes			
Immediate incentives	0.075**	0.057	0.093**	0.065	0.013	0.133**
	(0.037)	(0.064)	(0.044)	(0.055)	(0.034)	(0.057)
Observations	4,256	949	1,619	1,550	2,327	1,928
R-squared	0.129	0.154	0.151	0.132	0.125	0.198
Mean of Dep. Var.	3.499	3.465	3.491	3.495	3.694	3.263
Std. dev.	0.766	0.813	0.746	0.775	0.576	0.890
Number of individuals	825	600	765	744	786	688
			Panel B: Lectures			
Immediate incentives	0.089**	0.023	0.114**	0.096	0.030	0.133**
	(0.042)	(0.072)	(0.048)	(0.065)	(0.040)	(0.059)
Observations	3,382	728	1,371	1,219	1,700	1,681
R-squared	0.137	0.177	0.160	0.139	0.138	0.199
Mean of Dep. Var.	3.434	3.397	3.445	3.425	3.666	3.199
Std. dev.	0.807	0.861	0.777	0.817	0.600	0.914
Number of individuals	819	518	724	691	702	686

*Notes:*All estimates include demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), baseline sleep, indicators for the number of classes starting before 10am, indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. All estimates include controls for the time of the class. Columns 1-4 also include controls for whether the class is a STEM class or not. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline.

Turning to heterogeneity, Table 6 shows that the results are not driven by performance in early-morning classes but rather the largest effects are in late morning/early afternoon classes that occur between 10am and 2pm, followed by afternoon/evening classes (after 2pm). This is in line with the findings from Carrell et al. (2011), who find that early class start time affect performance in all classes, not just classes taking place early in the morning. Further, these results are driven by STEM courses: on average our intervention leads to a .133 grade points increased on grades in STEM courses, which corresponds to a 0.15 SD increase in grades. By contrast, we estimate small increases of 0.013 grade points in non-STEM courses. This finding suggests that sleep may have a more significant impact on quantitative courses. However, given the large proportion of STEM majors in our sample

(over 57%), another possibility is that improved sleep could enhance performance in courses important for a given major. In exploratory analysis, we find similar impacts by gender and evidence of particularly large effects among first-term freshmen (Appendix Table A.6).

To benchmark our intervention, we compare our effects to casual estimates of the relationship between sleep and academic performance from naturally occurring data. As discussed above, prior work examines the effect of shifts in sunset and school start times on sleep, grades and test scores. For example, Carrell et al. (2011) estimates that shifting the start time of college students' first class by an hour from 7:00 am to 8:05 am improves overall academic performance by 0.12 - 0.14 SD. The study does not directly measure students' sleep. Other studies using self-reported sleep estimate that an hour later school start time increases sleep by 35 minutes among American children with a 0.16 SD improvement in reading and no change in math (Groen and Pabilonia, 2019). Related studies find that the the sun rising one hour later increases average sleep among American children by an estimated six minutes with a 0.081 SD increase in math scores and a non-significant 0.057 SD improvement in reading scores (Heissel and Norris, 2018). Taken together, these studies suggest that a one hour shift increases sleep by 6 - 35 minutes and has either a null effect on academic performance or improves grades and test scores by 0.06 - 0.16 SD.[25] Our impacts of 19 minute average increase in weeknight sleep during treatment, an 9 minute average increases in post-treatment and a 0.10 - 0.11 SD improvement in grades falls within the range of the estimates in prior work on the causal relationship between shifts in sleep and changes in academic performance.

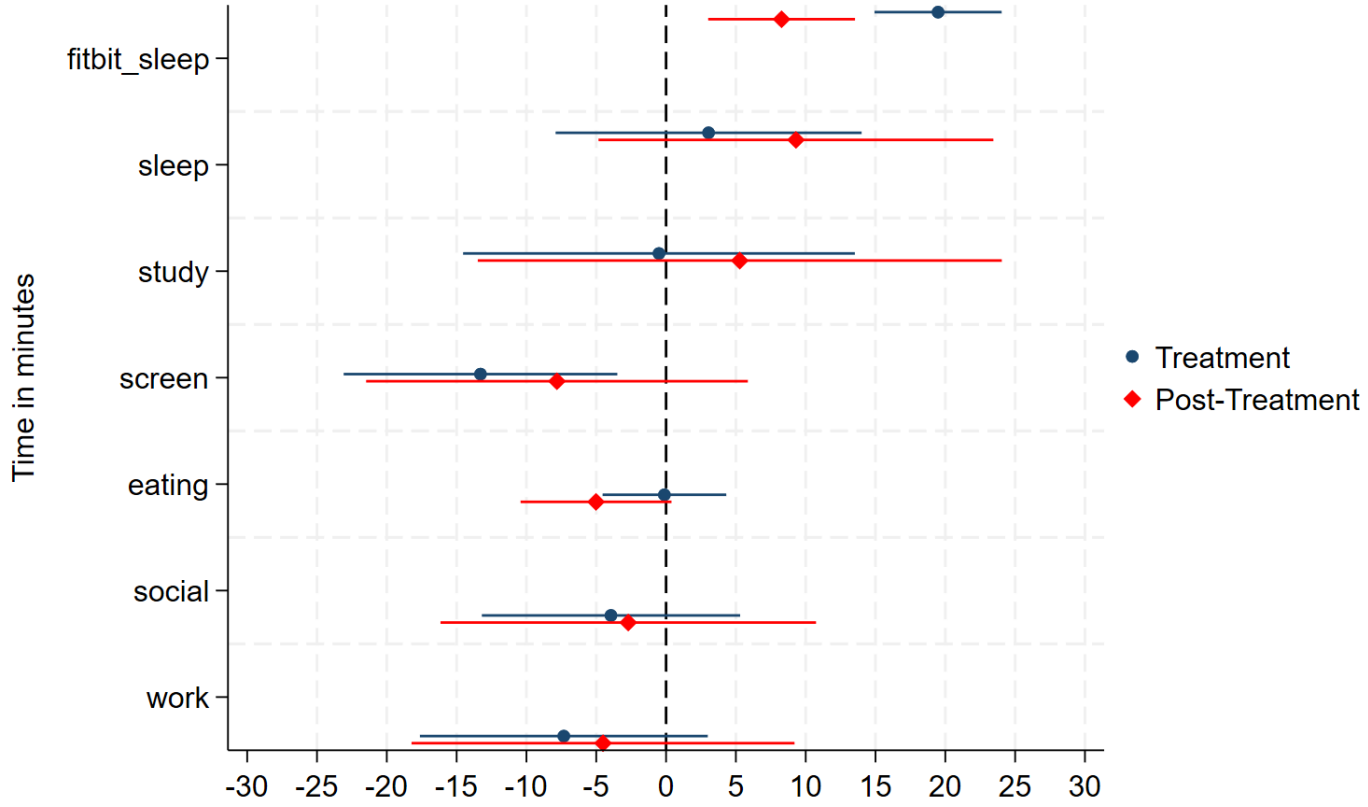
3.3 Additional Measures

To help make sense of our results, throughout the study we collected measures of time use, cognitive performance in math and creativity tasks, physical health via the Fitbit and well-being.

Lifestyle. We next focus on our survey measures of time use that we asked weekly throughout the study (see Section 2 for details). Figure 4 shows estimated treatment and post-treatment effects on time spent (in minutes) for the top six time use categories, using the regression specification in equation 1. In the first row of the figure, we show the estimated effects from the Fitbit data that we report in column 2 of Table 3. As discussed earlier, our intervention increased sleep by 19 minutes during treatment and 9 minutes after the removal

[25] Outside the U.S., Lusher et al. (2019) estimates that shifting class start times by an hour increases average sleep by about four minutes among Vietnamese University students with no effect on performance (Lusher et al., 2019). Jagnani (2021) estimates that the sun setting one hour earlier increases sleep by an average of 30 minutes among Indian children and that the sun setting 10 minutes earlier improves test scores by 0.1 SD and leads to 0.14 more years of schooling.

Figure 4: Incentives to sleep and time use (minutes)



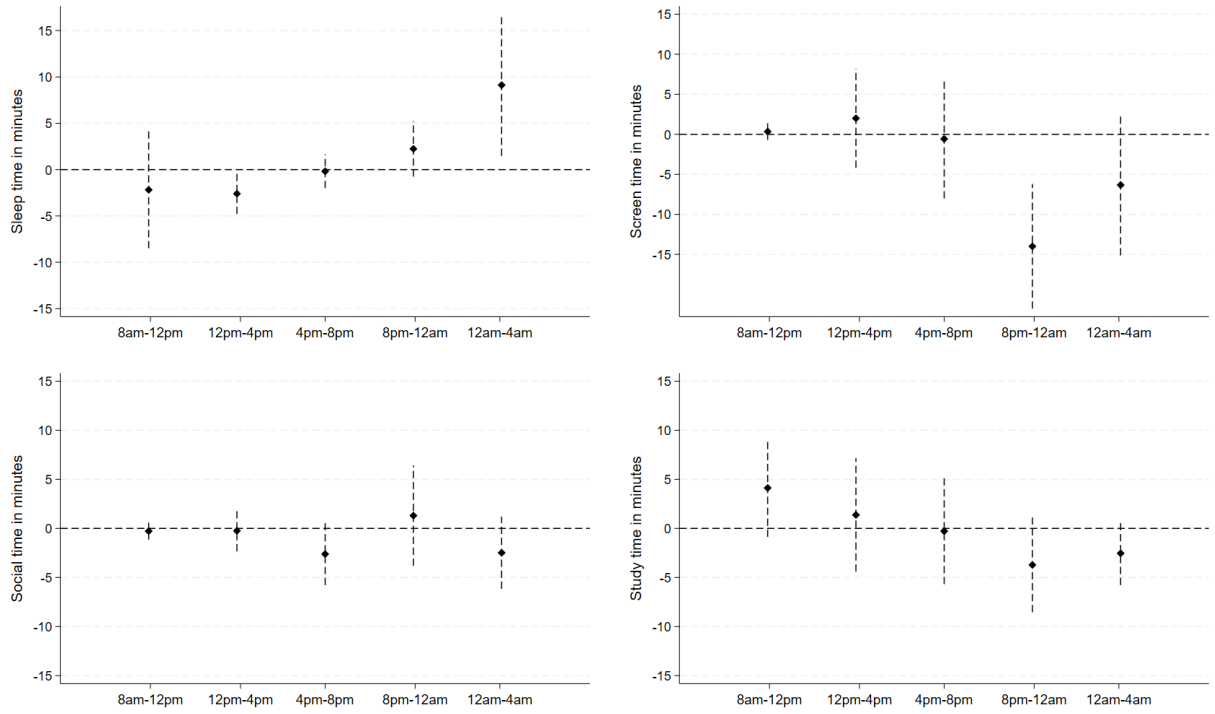
Notes - The figure reports differences between treatment and control in time-use during the intervention (navy) and in the post-intervention period (in red). All the coefficients are obtained from regressions including controls for day of the week fixed effects, month fixed effects, gender, race (dummies for Asian, Black, Hispanic, other) and ethnicity, indicators for the number of classes starting before 10am, parental education (dummies for less than college, college degree, and post-college degree), and high-school GPA. Standard errors are clustered at the individual level. Bars indicate 95% Confidence Intervals.

of the incentives. Immediate incentives directionally increase self-reported sleep in both the intervention and post-intervention periods by about 4 - 11 minutes on average per day.

During the intervention period, incentives to sleep significantly decrease average screen time, which includes internet browsing, TV/videos and games, excluding screen time for studying, by an estimated 14 minutes per day ($p < .01$). We estimate smaller and not statistically significant treatment effects on screen time during the post-intervention period. These estimates are robust to adjusting for multiple hypothesis testing ($p < .01$ for self-reported sleep during the intervention period; $p = .047$ for screen time). We do not find evidence of meaningful changes in time spent studying, socializing, eating or working.[26]

[26]We estimate treatment effects separately for internet, TV/videos and games in Appendix Table A.9 and the overall impact is largely driven by decreases in TV/video time. The table also reports effects on other

Figure 5: Immediate incentives to sleep and time use over the day: Intervention period



Notes - The figure reports differences between participants in treatment and control in the minutes allocated to different time-use activities during the intervention throughout the day. All the coefficients are obtained from regressions including controls for day of the week fixed effects, indicators for the number of classes starting before 10am, month fixed effects, gender, race (dummies for Asian, Black, Hispanic, other) and ethnicity, parental education (dummies for less than college, college degree, and post-college degree), and high-school GPA. Standard errors are clustered at the individual level. Bars indicate 95% Confidence Intervals.

In Figure 5, we report treatment effects on sleep, screen time, social time and study time over the course of the day during the intervention period. The effects on sleep and screen time are concentrated at night (8pm - 4am). Interestingly, although total study time does not increase, we observe a reallocation of study time from the evening/night (8pm—4am) to the morning (8am - 12pm), although the results are not precisely estimated. These results suggest that incentives to sleep led participants to develop sleep habits characterized by earlier screen disengagement at night and more focus on study time during the day. We also estimate treatment effects during the post-intervention period and find a similar, but weaker, patterns (Appendix A.4).

time use categories. At baseline, we estimate the following average minutes per day for each category: sleep (470 minutes), study (334 minutes), screen (157 minutes), eating and preparing food (90 minutes), social (95 minutes), work (88 minutes).

Cognitive Performance. To examine cognitive performance directly, we collected measures of performance in a math and creativity tasks on alternating weeks throughout the study. We do not find any impact of the intervention on these proxies for cognitive performance (see Appendix Table A.10, columns 1 and 2), which could be due to the intervention not affecting cognitive performance or to our measures not being able to capture the impact of performance on cognition.

Well-Being and Physical Health.

Table 7: Immediate Incentives and Well Being

	(1) Happiness	(2) Stress	(3) Resilience	(4) CES-D	(5) GAD-7
Treatment	-0.0687	0.0642	0.1514***		
	(0.108)	(0.059)	(0.055)		
Post-Treatment	-0.0651	-0.0375	0.0631	0.3854	0.0687
	(0.111)	(0.069)	(0.063)	(0.894)	(0.414)
Observations	4,166	3,629	3,558	1,462	1,462
R-squared	0.048	0.034	0.025	0.122	0.093
Mean of Dep. Var.	6.404	3.115	2.993	15.78	6.832
Std. dev.	1.646	1.116	0.997	10.23	4.864
Number of individuals	794	800	794	834	834

Notes:
All estimates include day of the week, week of the experiment, wave, and month fixed effects, baseline value of the outcome variable, and demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), indicators for the number of classes starting before 10am,indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. For mood, participants indicated, on a 10-point Likert scale, how happy they felt in that moment. For stress and resilience, participants indicated, using a 5-point Likert scale, 1) the extent to which participants faced stress in their life at the time of answering the survey and 2) the extent to which they felt able to deal with the stress they were facing. CES-D is the Center for Epidemiologic Studies Depression Scale. GAD-7 is the General Anxiety Disorder-7. Standard errors are clustered at the individual level. Mean of Dep. Var. is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline. *** p<0.01, ** p<0.05, * p<0.1

Our final outcomes of interest are well-being and physical health. Previous work suggests that there is a positive relationship between sleep and both mental wellbeing and physical health (Giuntella and Mazzonna, 2019; Giuntella et al., 2017; Jin and Ziebarth, 2020).

To investigate the impact of the intervention on well-being by employing weekly text messages to collect data on mood, stress, and resilience to stress. Additionally, we utilize the Generalized Anxiety Disorder (GAD-7) scale to assess anxiety levels and the Center for Epidemiologic Studies Depression (CES-D) scale to gauge depression levels. These scales were administered at baseline and endline only and so we are only able to estimate treatment effects on post-intervention end-of-semester anxiety and depression. Table 7 shows that the intervention does not have a significant impact on mood or stress levels (columns 1 and 2). However, it led to a statistically significant increase in resilience—participants' self-reported

ability to cope with stress—by approximately 0.15 standard deviations (column 3), which is significant at the 10% level after adjusting for multiple hypothesis testing. On the other hand, the intervention did not show any significant effects on post-treatment measures of depression and anxiety (columns 4 and 5), with point estimates being small in magnitude and lacking statistical significance.

As discussed in Section 2, we use the Fitbit to measure participants' heart rate, daily steps, physical activity. We present estimates of treatment and post-treatment effects in Table A.10. We find no evidence of treatment effects on any of the physical health measures (columns 3 and 5).

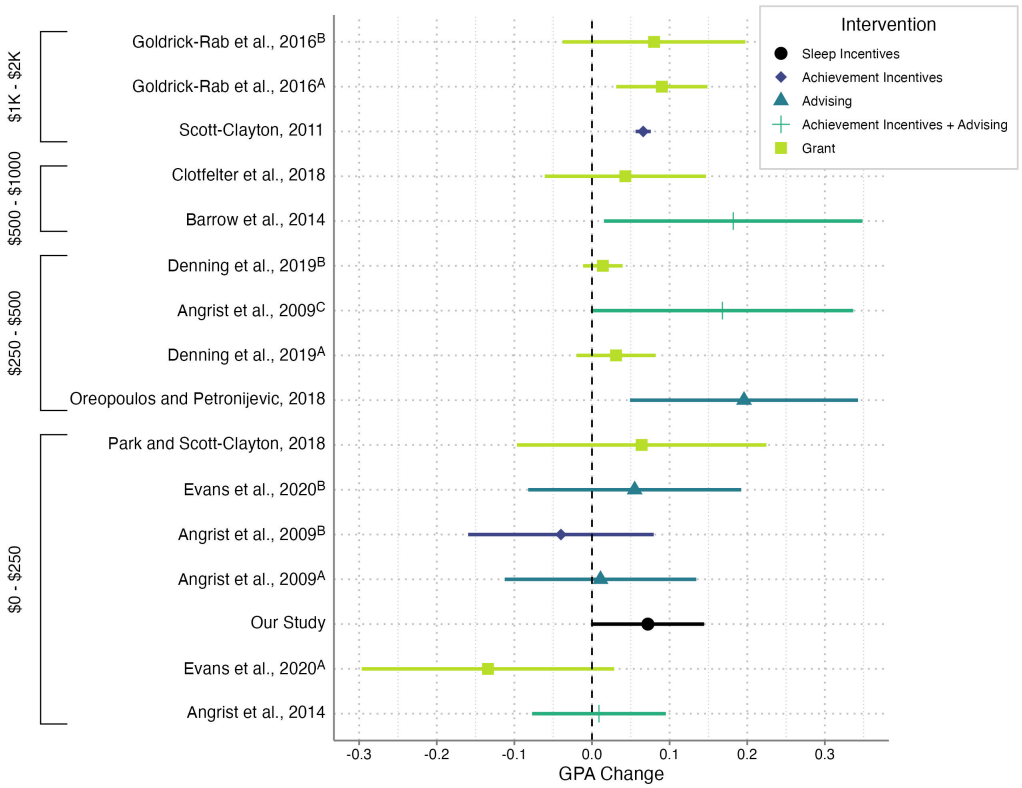
4 Discussion and Conclusion

In this paper, we show that an intervention targeting sleep habits improves academic performance. To benchmark our results, we compare the cost effectiveness of our intervention to prior work examining policies aimed at improve college students' outcomes, including those that condition rewards on academic performance. Angrist et al. (2014) summarize the work on performance based incentives, including their own, and conclude that, "the picture that emerges. . . is one of mostly modest effects. . . [And there are] similarly discouraging results from studies of state-based merit aid programs. A few studies report positive effects, most notably Scott-Clayton (2011)'s evaluation of West Virginia PROMISE," which conditions free tuition on meeting a minimum GPA. Scott-Clayton (2011) finds similar-sized GPA effects to our study, 0.066 grade point improvements, but at over ten times the cost, an estimated $1250 per student per semester. By comparison, we estimate that incentives to sleep increase semester GPA by 0.075 grade points and cost approximately $110 per participant for the semester and would cost about $160 per participant per year. This includes $60 for the cost of the Fitbit and an estimated average of $52 per participant per semester for the incentives (participants in the immediate incentives group received the incentives of $4.75 per night on 55% of the 20 nights we offered it).

In Figure 6, we report the estimated effects on GPA from prior work examining the impact of achievement incentives, advising, and grants, ordered from most to least costly.[27] As depicted in the figure, our intervention is characterized by relatively low costs, while the estimated effects are equal to or greater than those observed in most previous studies. Only

[27] Achievement incentives include performance-based incentives and merit aid. Advising includes advising and support services, see A.11. We note that for some of these programs the primary outcome may have been enrollment, persistence or graduation and GPA may have been a secondary outcome.

Figure 6: Cost-Effectiveness



Notes - The figure compares our main effects on GPA to estimates from previous interventions aimed at improving college academic performance. Studies are grouped on the vertical axis based on their cost per subject per semester. Bars represent 95% confidence intervals. Superscripts above paper names denote different treatment arms or treatment groups. For Goldrick-Rab et al. (2016), superscript A is an estimate for the first cohort studied and B is their pooled estimate for the second and third cohort. For Denning et al. (2019), A and B are estimates for first-year and returning students, respectively. For Angrist et al. (2009), A is an estimate for an advising and peer-support treatment arm, B is for a financial incentives arm, and C for an arm combing A and B. For Evans et al. (2020), A estimates a grant treatment arm, and B estimates combined grant aid with academic advising.

a handful of interventions surpass ours in terms of impact, but they come with a two to fivefold higher cost per participant. Our results demonstrate that focusing on sleep can be a cost-effective approach to improving educational outcomes.

We explore mechanisms for the impact of our intervention on both sleep habits and academic performance. Inspired by cue-based theories of habit formation, our immediate incentives intervention aimed to establish automatic habits through repeated exposure to recurring cues coupled with immediate rewards. The intervention increases sleep during the treatment period with smaller persistent effects in the post-treatment period. Our results show that immediate incentives can enhance habit formation during the intervention period, compared to variants with delayed or no rewards. However, we find little evidence that immediate incentives generate automatic habits triggered by the external cue. Instead,

our results point to participants developing their own routines that persist into the post-treatment period. This could reflect some combination of treated participants acquiring a taste for sleep (i.e., increased benefits) and also finding sleep behaviors that are easier to sustain (i.e., lower costs). Future research could develop intervention designs that separately identify mechanisms of habit formation, including, automaticity, learning about benefits and lowering costs (Volpp and Loewenstein, 2020).

We then examine channels through which sleep may influence academic performance, including cognitive function, lifestyle factors, and overall well-being. While we do not detect an impact of our intervention on performance in math questions or creativity, sleep could have influenced cognition through channels like attention or memory consolidation, which were not captured by our measures (Diekelmann and Born, 2010), which we did not capture in our study. Examining lifestyle, our intervention led to a decrease in screen time and a reallocation of study time to morning hours, when students are potentially more alert and able to focus. Finally, we find evidence of a positive impact on students' ability to cope with stress, which may in turn have affected their academic performance. Further investigation of these mechanisms in future research can provide a deeper understanding of the multifaceted contributions of sleep to educational outcomes.

Taken together, our results show that offering incentives in the middle of the semester can improve term GPA. Future work could examine targeting the intervention (for example to first-term freshmen), the role of the timing of the intervention, and the impact of longer (or shorter) interventions in order to understand how to cost-effectively sustain effects on academic performance across multiple terms.

References

Abraham, K. G. and Flood, S. M. (2009). American time use survey data extract builder (atus-x). *International Journal for Time Use Research*, 6:167–168.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Angrist, J., Lang, D., and Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1):136–163.

Angrist, J., Oreopoulos, P., and Williams, T. (2014). When opportunity knocks, who answers? new evidence on college achievement awards. *Journal of Human Resources*, 49(3):572–610.

Avery, M., Giuntella, O., and Jiao, P. (2022). Why don't we sleep enough? a field experiment among college students. *Review of Economics and Statistics, Forthcoming*.

Bachireddy, C., Joung, A., John, L. K., Gino, F., Tuckfield, B., Foschini, L., and Milkman, K. L. (2019). Effect of different financial incentive structures on promoting physical activity among adults: A randomized clinical trial. *JAMA network open*, 2(8):e199863–e199863.

Ballard, J. (2019). 40% of americans don't generally wake up feeling well-rested. *YouGov*.

Banks, S. and Dinges, D. F. (2007). Behavioral and physiological consequences of sleep restriction. *Journal of clinical sleep medicine*, 3(5):519–528.

Barnes, C. M., Miller, J. A., and Bostock, S. (2017). Helping employees sleep well: Effects of cognitive behavioral therapy for insomnia on work outcomes. *Journal of Applied Psychology*, 102(1):104.

Barrow, L., Richburg-Hayes, L., Rouse, C. E., and Brock, T. (2014). Paying for performance: The education impacts of a community college scholarship program for low-income adults. *Journal of Labor Economics*, 32(3):563–599.

Basner, M., Fomberstein, K. M., Razavi, F. M., Banks, S., William, J. H., Rosa, R. R., and Dinges, D. F. (2007). American time use survey: sleep time and its relationship to waking activities. *Sleep*, 30(9):1085–1095.

Beshears, J., Lee, H. N., Milkman, K. L., Mislavsky, R., and Wisdom, J. (2021). Creating exercise habits using incentives: The trade-off between flexibility and routinization. *Management science*, 67(7):4139–4171.

Bessone, P., Rao, G., Schilbach, F., Schofield, H., and Toma, M. (2021). The economic consequences of increasing sleep among the urban poor. *The Quarterly Journal of Economics*, 136(3):1887–1941.

Biddle, J. E. and Hamermesh, D. S. (1990). Sleep and the allocation of time. *Journal of Political Economy*, 98(5, Part 1):922–943.

Boyce, R., Glasgow, S. D., Williams, S., and Adamantidis, A. (2016). Causal evidence for the role of rem sleep theta rhythm in contextual memory consolidation. *Science*, 352(6287):812–816.

Byrne, D. P., Goette, L., Martin, L. A., Miles, A., Jones, A., Schob, S., Staake, T., and Tiefenbeck, V. (2022). The habit forming effects of feedback: Evidence from a large-scale field experiment. *Available at SSRN 3974371*.

Cappelen, A. W., Charness, G., Ekström, M., Gneezy, U., and Tungodden, B. (2017). Exercise improves academic performance. *NHH Dept. of Economics Discussion Paper*, (08).

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

Charness, G. and Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77(3):909–931.

Charness, G. and Grieco, D. (2019). Creativity and incentives. *Journal of the European Economic Association*, 17(2):454–496.

Clotfelter, C. T., Hemelt, S. W., and Ladd, H. F. (2018). Multifaceted aid for low-income students and college outcomes: Evidence from north carolina. *Economic Inquiry*, 56(1):278–303.

Cooney, M. T., Vartiainen, E., Laakitainen, T., Juolevi, A., Dudina, A., and Graham, I. M. (2010). Elevated resting heart rate is an independent risk factor for cardiovascular disease in healthy men and women. *American heart journal*, 159(4):612–619.

Creswell, J. D., Tumminia, M. J., Price, S., Sefidgar, Y., Cohen, S., Ren, Y., Brown, J., Dey, A. K., Dutcher, J. M., Villalba, D., et al. (2023). Nightly sleep duration predicts grade point average in the first year of college. *Proceedings of the National Academy of Sciences*, 120(8):e2209123120.

de Zambotti, M., Goldstone, A., Claudatos, S., Colrain, I. M., and Baker, F. C. (2018). A validation study of fitbit charge 2™ compared with polysomnography in adults. *Chronobiology international*, 35(4):465–476.

Denning, J. T., Marx, B. M., and Turner, L. J. (2019). Propelled: The effects of grants on graduation, earnings, and welfare. *American Economic Journal: Applied Economics*, 11(3):193–224.

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1135):67–78.

Dickinson, D. L. and Masclet, D. (2021). Unethical decision making and sleep restriction: Experimental evidence.

Dickinson, D. L. and McElroy, T. (2017). Sleep restriction and circadian effects on social decisions. *European Economic Review*, 97:57–71.

Diekelmann, S. and Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2):114–126.

Dyer, A. R., Persky, V., Stamler, J., Paul, O., Shekelle, R. B., Berkson, D. M., Lepper, M., Schoenberger, J. A., and Lindberg, H. A. (1980). Heart rate as a prognostic factor for coronary heart disease and mortality: findings in three chicago epidemiologic studies. *American journal of epidemiology*, 112(6):736–749.

Evans, W. N., Kearney, M. S., Perry, B., and Sullivan, J. X. (2020). Increasing community college completion rates among low-income students: Evidence from a randomized controlled trial evaluation of a case-management intervention. *Journal of Policy Analysis and Management*, 39(4):930–965.

Gibson, M. and Shrader, J. (2018). Time use and labor productivity: The returns to sleep. *Review of Economics and Statistics*, 100(5):783–798.

Giuntella, O., Han, W., and Mazzonna, F. (2017). Circadian rhythms, sleep, and cognitive skills: Evidence from an unsleeping giant. *Demography*, 54(5):1715–1742.

Giuntella, O., Hyde, K., Saccardo, S., and Sadoff, S. (2021). Lifestyle and mental health disruptions during covid-19. *Proceedings of the National Academy of Sciences*, 118(9):e2016632118.

Giuntella, O. and Mazzonna, F. (2019). Sunset time and the economic effects of social jetlag: evidence from us time zone borders. *Journal of health economics*, 65:210–226.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of economic perspectives*, 25(4):191–210.

Goldrick-Rab, S., Kelchen, R., Harris, D. N., and Benson, J. (2016). Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion. *American Journal of Sociology*, 121(6):1762–1817.

Groen, J. A. and Pabilonia, S. W. (2019). Snooze or lose: High school start times and academic achievement. *Economics of Education Review*, 72:204–218.

Hafner, M., Stepanek, M., Taylor, J., Troxel, W. M., and Van Stolk, C. (2017). Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis. *Rand health quarterly*, 6(4).

Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R., and Castriotta, R. J. (2019). Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis. *Journal of medical Internet research*, 21(11):e16273.

Heissel, J. A. and Norris, S. (2018). Rise and shine the effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources*, 53(4):957–992.

Hershner, S. D. and Chervin, R. D. (2014). Causes and consequences of sleepiness among college students. *Nature and science of sleep*, pages 73–84.

Hillman, D. R., Murphy, A. S., Antic, R., and Pezzullo, L. (2006). The economic cost of sleep disorders. *Sleep*, 29(3):299–305.

Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Hillard, P. J. A., Katz, E. S., et al. (2015). National sleep foundation's updated sleep duration recommendations. *Sleep health*, 1(4):233–243.

Holbein, J. B., Schafer, J. P., and Dickinson, D. L. (2019). Insufficient sleep reduces voting and other prosocial behaviours. *Nature human behaviour*, 3(5):492–500.

Hull, C. L. (1943). Principles of behavior: an introduction to behavior theory.

Hussam, R. N., Rabbani, A., Reggiani, G., and Rigol, N. (2017). Habit formation and rational addiction: A field experiment in handwashing. *Harvard Business School working paper series# 18-030*.

Jagnani, M. (2021). Children's sleep and human capital production. *The Review of Economics and Statistics*, pages 1–45.

Jike, M., Itani, O., Watanabe, N., Buysse, D. J., and Kaneita, Y. (2018). Long sleep duration and health outcomes: a systematic review, meta-analysis and meta-regression. *Sleep medicine reviews*, 39:25–36.

Jin, L. and Ziebarth, N. R. (2020). Sleep, health, and human capital: Evidence from daylight saving time. *Journal of Economic Behavior & Organization*, 170:174–192.

Jones, J. (2013). In u.s., 40% get less than recommended amount of sleep. *Gallup*.

Killgore, W. D. (2010). Effects of sleep deprivation on cognition. *Progress in brain research*, 185:105–129.

Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K., and Rathouz, P. J. (2008). Self-reported and measured sleep duration: how similar are they? *Epidemiology*, pages 838–845.

Leproult, R. and Van Cauter, E. (2010). Role of sleep and sleep loss in hormonal release and metabolism. *Pediatric Neuroendocrinology*, 17:11–21.

Levitt, S. D., List, J. A., Neckermann, S., and Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4):183–219.

Lindquist, S. and Sadoff, S. (2023). Understanding the interactions of sleep, social media and mental health for productivity and performance: The role of field experiments. *Annual Proceedings of the Upton Forum*.

Lusher, L., Yasenov, V., and Luong, P. (2019). Does schedule irregularity affect productivity? evidence from random assignment into college classes. *Labour Economics*, 60:115–128.

Marks, G. A., Shaffery, J. P., Oksenberg, A., Speciale, S. G., and Roffwarg, H. P. (1995). A functional role for rem sleep in brain maturation. *Behavioural brain research*, 69(1-2):1–11.

McKenna, B. S., Dickinson, D. L., Orff, H. J., and Drummond, S. P. (2007). The effects of one night of sleep deprivation on known-risk and ambiguous-risk decisions. *Journal of sleep research*, 16(3):245–252.

Milkman, K. L., Minson, J. A., and Volpp, K. G. (2014). Holding the hunger games hostage at the gym: An evaluation of temptation bundling. *Management science*, 60(2):283–299.

Mullainathan, S. (2014). Get some sleep, and wake up the g.d.p. *The New York Times*.

Oreopoulos, P. and Petronijevic, U. (2018). Student coaching: How far can technology go? *Journal of Human Resources*, 53(2):299–329.

Panel, C. C., Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., Dinges, D. F., Gangwisch, J., Grandner, M. A., et al. (2015). Recommended amount of sleep for a healthy adult: a joint consensus statement of the american academy of sleep medicine and sleep research society. *Journal of Clinical Sleep Medicine*, 11(6):591–592.

Park, R. S. E. and Scott-Clayton, J. (2018). The impact of pell grant eligibility on community college students' financial aid packages, labor supply, and academic outcomes. *Educational Evaluation and Policy Analysis*, 40(4):557–585.

Phillips, A. J., Clerx, W. M., O'Brien, C. S., Sano, A., Barger, L. K., Picard, R. W., Lockley, S. W., Klerman, E. B., and Czeisler, C. A. (2017). Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Scientific reports*, 7(1):3216.

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401.

Roenneberg, T. (2013). The human sleep project. *Nature*, 498(7455):427–428.

Royer, H., Stehr, M., and Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, 7(3):51–84.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434.

Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. *Journal of Human resources*, 46(3):614–646.

Smarr, B. L. (2015). Digital sleep logs reveal potential impacts of modern temporal structure on class performance in different chronotypes. *Journal of biological rhythms*, 30(1):61–67.

Trockel, M. T., Barnes, M. D., and Egget, D. L. (2000). Health-related variables and academic performance among first-year college students: Implications for sleep and other behaviors. *Journal of American college health*, 49(3):125–131.

Verplanken, B. (2006). Beyond frequency: Habit as mental construct. *British Journal of Social Psychology*, 45(3):639–656.

Verplanken, B. and Wood, W. (2006). Interventions to break and create consumer habits. *Journal of public policy & marketing*, 25(1):90–103.

Volpp, K. G. and Loewenstein, G. (2020). What is a habit? diverse mechanisms that can produce sustained behavior change. *Organizational Behavior and Human Decision Processes*, 161:36–38.

Walker, M. (2017). *Why we sleep: Unlocking the power of sleep and dreams*. Simon and Schuster.

Wellsjo, A. S. (2021). Simple actions, complex habits: Lessons from hospital hand hygiene. Technical report, Working paper.

Williams, N. (2014). The gad-7 questionnaire. *Occupational medicine*, 64(3):224–224.

Wood, W. and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological review*, 114(4):843.

Wood, W. and Rünger, D. (2016). Psychology of habit. *Annual review of psychology*, 67:289–314.

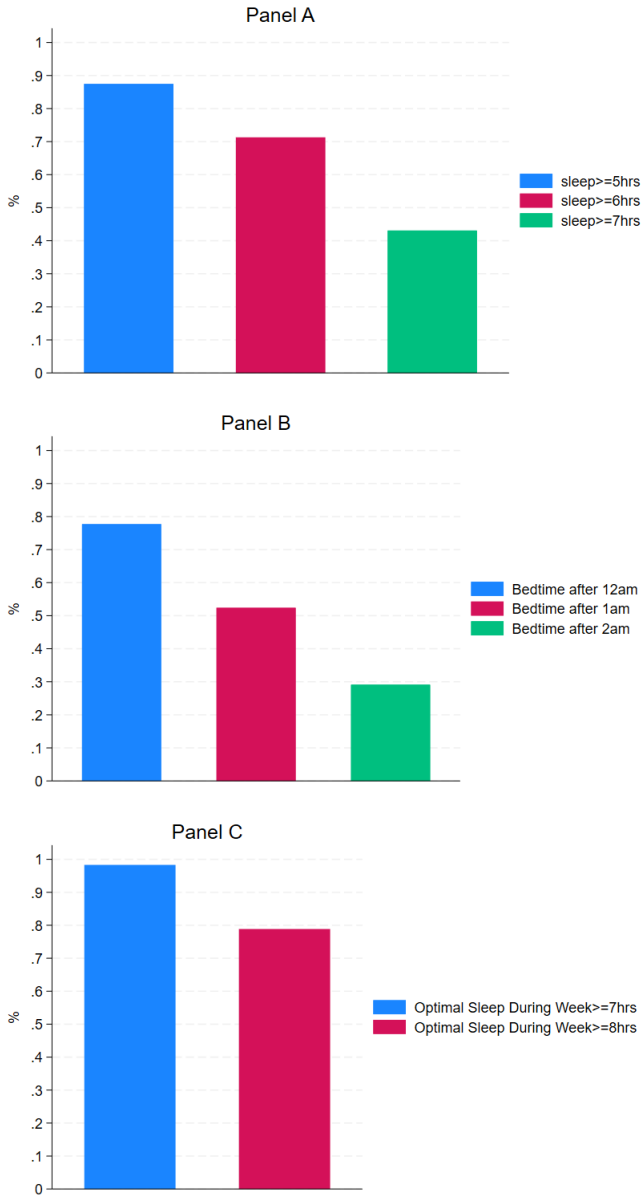
Wright, S. P., Hall Brown, T. S., Collier, S. R., and Sandberg, K. (2017). How consumer physical activity monitors could transform human physiology research. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 312(3):R358–R367.

Zhang, Y. and Gruber, R. (2019). Focus: attention science: can slow-wave sleep enhancement improve memory? a review of current approaches and cognitive outcomes. *The Yale journal of biology and medicine*, 92(1):63.

Appendix

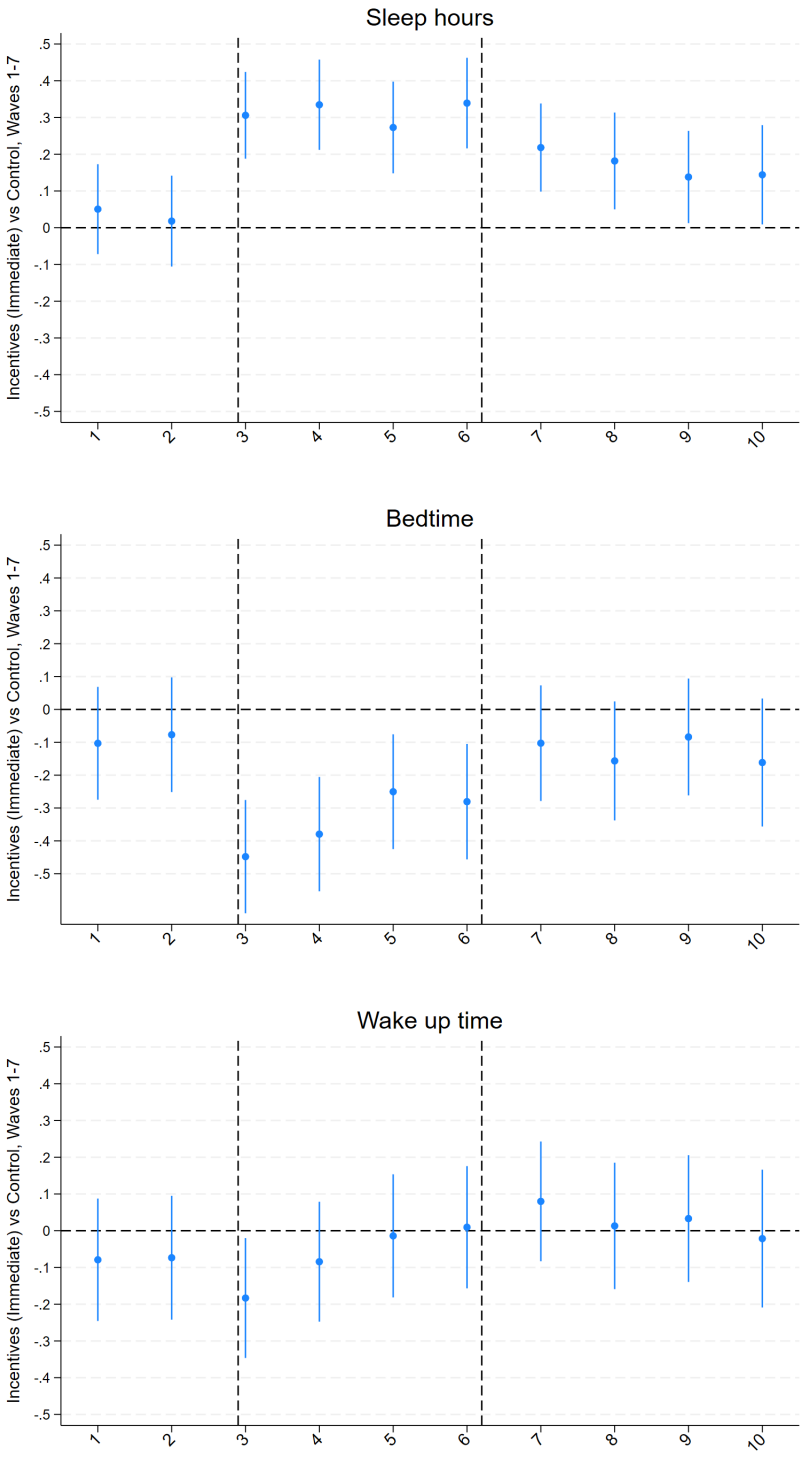
A. Figures and Tables

Figure A.1: Sleep, Bedtime, and Optimal Week Sleep at Baseline



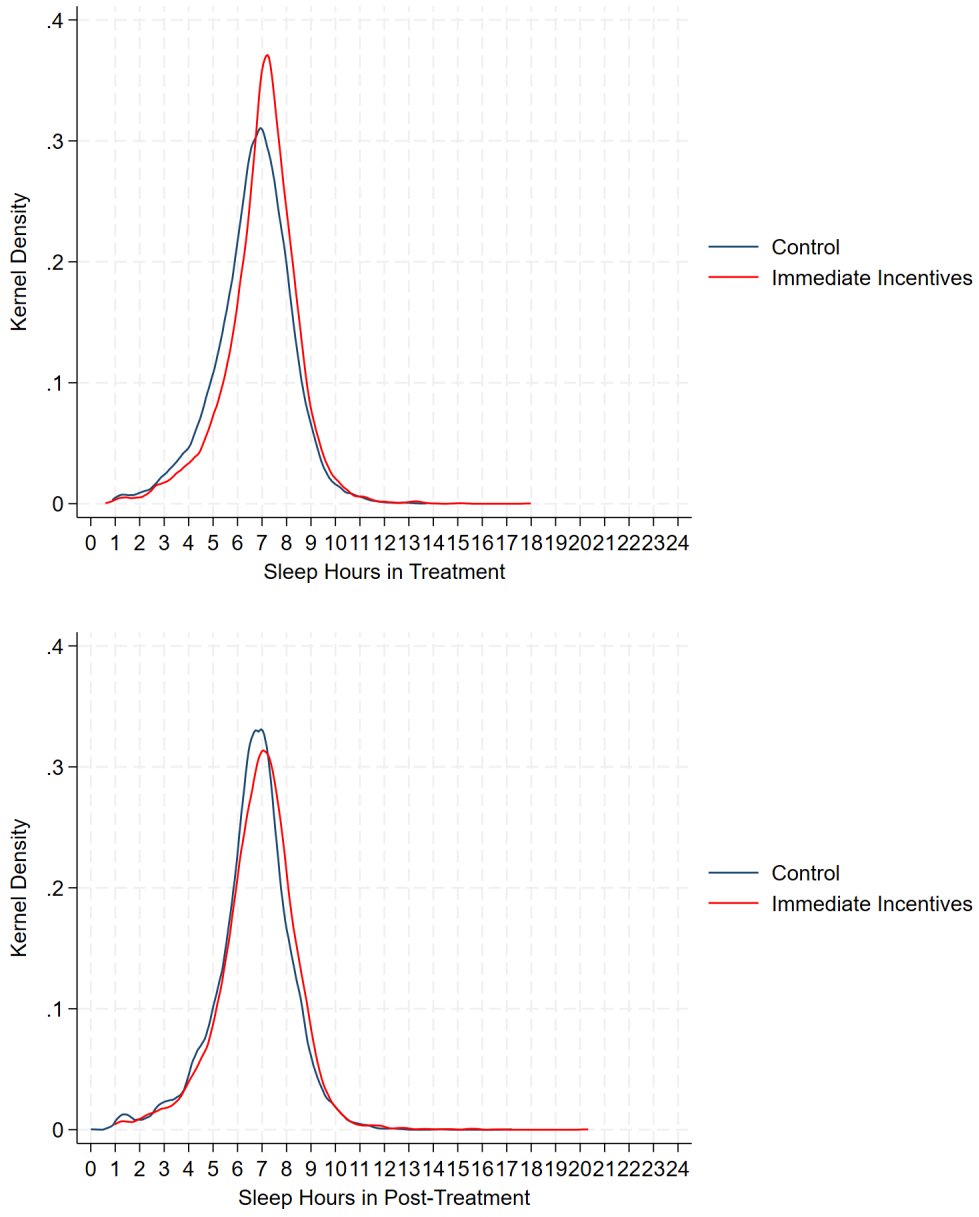
Notes: Panel A reports the share of nights with at least 7 hours of sleep (green), 6 hours (red), and 5 hour (blue) at baseline as measured by Fitbits. Panel B reports the share of participants going to bed after midnight (blue), after 1am (red), and after 2am (green) at baseline as measured by Fitbits. Panel C reports the share of participants reporting that their optimal sleep during the week would be at least 7 hours (blue), 8 hours (red) in the intake survey.

Figure A.2: Immediate Incentives, Sleep Hours, Bedtime and Wake-up Time



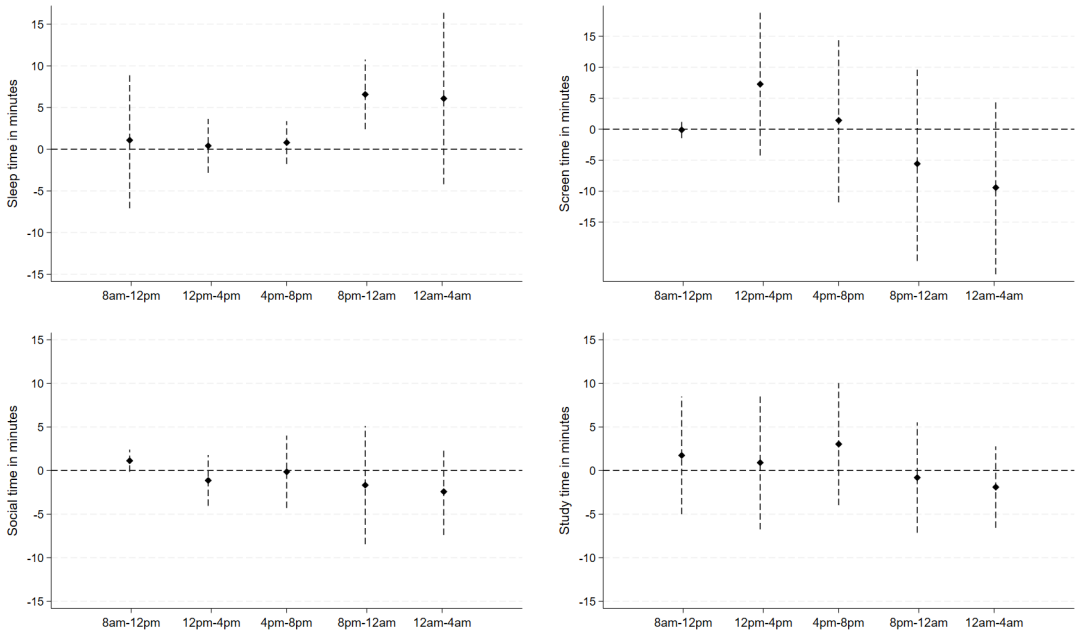
Notes - The sample is restricted to weekdays (Sunday-Thursday nights). On the horizontal axis we report time in weeks since the study started (week=3 is the first week of treatment, week=6 is the last week of treatment). The coefficient reports the differences in average sleep hours, bedtime, and wake up time between individuals in the immediate incentives treatment and those in control by week. Standard errors are clustered at the individual level. Bars indicate 95% Confidence Intervals.

Figure A.3: Immediate incentives and distribution of sleep during and after the intervention



Notes: The figure reports the distribution of sleep hours during (top figure) and after (bottom figure) the intervention for participants in control (navy) and treatment (red).

Figure A.4: Immediate incentives to sleep and time use over the day: Post-Intervention period



Notes - The figure reports differences between participants in treatment and control in the minutes allocated to different time-use activities during the intervention throughout the day. All the coefficients are obtained from regressions including controls for day of the week fixed effects, month fixed effects, indicators for the number of classes starting before 10am, gender, race (dummies for Asian, Black, Hispanic, other) and ethnicity, parental education (dummies for less than college, college degree, and post-college degree), and high-school GPA. Standard errors are clustered at the individual level. Bars indicate 95% Confidence Intervals.

Table A.1: Incentives and Attrition Across Surveys

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A			
	# Fitbit days before intervention	# Fitbit days during intervention	# Fitbit days after intervention	Has HS GPA	Has Baseline GPA	Has Term GPA	Has Course Grades
Immediate incentives	0.106	0.780**	-0.069	0.017	0.045*	0.023	-0.005
	(0.257)	(0.320)	(0.407)	(0.020)	(0.026)	(0.016)	(0.003)
Observations	840	840	840	840	840	840	777
Mean of Dep. Var.	11.89	18.17	13.03	0.895	0.782	0.956	0.873
Std. dev.	6.021	4.100	6.428	0.307	0.413	0.205	0.333
				Panel B			
	Has time use	Has math	Has creativity	Has mood	Has resilience	GAD-7	
Immediate incentives	0.006	0.009	0.010	0.005	-0.003	-0.001	
	(0.007)	(0.013)	(0.015)	(0.011)	(0.012)	(0.006)	
Observations	840	840	840	840	840	840	
Mean of Dep. Var.	0.988	0.872	0.951	0.967	0.959	0.991	
Std. dev.	0.110	0.334	0.216	0.178	0.198	0.0931	

Notes: The table reports the difference between treatment and control group in attrition rate across the different surveys. Robust standard errors are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Differences in Baseline Characteristics: Secondary Treatments

	(1) Sleep ≥ 7 hours	(2) Sleep ≥ 6 hours	(3) Sleep hours	(4) Bedtime	(5) Wake up time	(6) HS GPA	(7) Baseline GPA	(8)
Delayed incentives	0.0081	-0.0338	-0.0197	-0.2647*	-0.3047**	0.0633	-3.2019	
	(0.120)	(0.034)	(0.029)	(0.146)	(0.142)	(0.053)	(5.360)	
Delayed incentives, no feedback	-0.0731	-0.0442	-0.0246	0.0056	-0.0501	0.0952*	-6.8703	
	(0.108)	(0.033)	(0.030)	(0.150)	(0.149)	(0.057)	(5.309)	
Feedback only	-0.0199	-0.0210	-0.0092	-0.2074	-0.2445*	0.1036*	-6.4238	
	(0.101)	(0.033)	(0.030)	(0.143)	(0.144)	(0.059)	(5.245)	
Observations	444	444	444	444	444	385	351	
R-squared	0.040	0.060	0.022	0.021	0.045	0.014	0.241	
Mean of Dep. Var.	6.754	0.447	0.733	24.99	7.774	4.126	25.96	
Std. dev.	0.841	0.259	0.230	1.128	1.116	0.399	40.71	
Immediate incentives	-0.0058	-0.0302	-0.1193	-0.0024	-0.0379	-0.0161	-0.1087*	
no cue/feedback in post	(0.036)	(0.034)	(0.120)	(0.175)	(0.162)	(0.060)	(0.061)	
Observations	224	224	224	224	224	215	217	
Mean of Dep. Var.	0.467	0.739	6.794	25.23	8.083	4.138	3.503	
Std. dev.	0.272	0.257	0.912	1.303	1.232	0.444	0.448	
	Freshman	Sophomore	Junior	Senior	STEM major	Less than College	College	More than College
Delayed incentives	-0.0242	-0.0466	0.0479	0.0354	0.1369**	-0.0274	0.0350	-0.0077
	(0.063)	(0.045)	(0.056)	(0.049)	(0.062)	(0.059)	(0.062)	(0.064)
Delayed incentives, no feedback	-0.0327	-0.0361	0.0218	0.0595	0.0051	0.0196	0.0565	-0.0760
	(0.065)	(0.048)	(0.056)	(0.052)	(0.067)	(0.061)	(0.063)	(0.065)
Feedback only	-0.0303	0.0075	-0.0164	0.0516	0.1016	-0.0123	-0.0069	0.0192
	(0.062)	(0.049)	(0.053)	(0.051)	(0.064)	(0.060)	(0.060)	(0.065)
Observations	444	444	444	444	442	442	442	442
R-squared	0.045	0.025	0.015	0.028	0.015	0.004	0.006	0.007
Mean of Dep. Var.	0.432	0.160	0.223	0.180	0.620	0.287	0.312	0.400
Std. dev.	0.496	0.367	0.417	0.385	0.486	0.453	0.464	0.491
Immediate incentives	0.1002*	-0.0253	-0.0412	-0.0337	0.0718	0.0321	0.1034*	-0.1355**
no cue/feedback in post	(0.060)	(0.037)	(0.051)	(0.042)	(0.067)	(0.062)	(0.061)	(0.066)
Observations	224	224	224	224	224	224	224	224
Mean of Dep. Var.	0.607	0.0848	0.192	0.116	0.522	0.299	0.290	0.411
Std. dev.	0.489	0.279	0.395	0.321	0.501	0.459	0.455	0.493
	Female	Age	Asian	Black	Hispanic	White	Other	
Delayed incentives	0.0111	0.1893	-0.0640	0.0126	-0.0104	0.0698	-0.0080	
	(0.060)	(0.492)	(0.059)	(0.037)	(0.021)	(0.065)	(0.022)	
Delayed incentives, no feedback	-0.0164	-0.0607	-0.0290	-0.0020	0.0110	0.0160	0.0041	
	(0.062)	(0.390)	(0.061)	(0.034)	(0.027)	(0.066)	(0.026)	
Feedback only	0.0051	-0.0994	-0.1195**	0.0636	0.0194	0.0143	0.0222	
	(0.060)	(0.385)	(0.056)	(0.041)	(0.028)	(0.066)	(0.027)	
Observations	442	442	442	442	442	442	442	
R-squared	0.045	0.038	0.018	0.024	0.010	0.010	0.004	
Mean of Dep. Var.	0.686	19.83	0.271	0.0973	0.0362	0.554	0.0407	
Std. dev.	0.465	3.270	0.445	0.297	0.187	0.498	0.198	
Immediate incentives	0.0139	-0.0302	0.0582	-0.0344	0.0190	-0.0083	-0.0344	
no cue/feedback in post	(0.056)	(0.252)	(0.058)	(0.034)	(0.033)	(0.066)	(0.034)	
Observations	224	224	224	224	224	224	224	
Mean of Dep. Var.	0.777	19.31	0.259	0.0714	0.0625	0.536	0.0714	
Std. dev.	0.417	1.927	0.439	0.258	0.243	0.500	0.258	

Notes: All estimates include wave fixed effects. Robust standard errors are in parenthesis.

Table A.3: Incentives and Sleep: Other Outcomes

	any nap	sleep ≥ 7 weekends	sleep ≥ 7 weekends & holidays	sleep ≥ 7 all nights & naps	sleep hours all nights & naps
Treatment	0.0023	0.0144	0.0099	0.0690***	0.2068***
	(0.004)	(0.015)	(0.016)	(0.011)	(0.032)
Post-Treatment	-0.0049	0.0491***	0.0467***	0.0415***	0.0957***
	(0.004)	(0.016)	(0.016)	(0.012)	(0.032)
Observations	69,937	18,100	22,948	69,937	69,937
R-squared	0.075	0.196	0.145	0.304	0.309
Mean of Dep. Var.	0.0513	0.478	0.478	0.568	7.213
Std. dev.	0.188	0.500	0.500	0.495	1.425
Number of individuals	840	840	840	840	840
	sleep ≥ 6	sleep 7-9	Efficiency	REM sleep	Deep sleep
Treatment	0.0801***	0.1098***	0.2429*	2.4392***	0.3557
	(0.010)	(0.011)	(0.144)	(0.775)	(0.651)
Post-Treatment	0.0368***	0.0494***	0.0624	1.7100*	0.0929
	(0.010)	(0.012)	(0.184)	(0.877)	(0.698)
Observations	46,989	46,989	46,989	43,168	43,168
R-squared	0.246	0.190	0.599	0.342	0.300
Mean of Dep. Var.	0.714	0.382	93.52	84.11	74.51
Std. dev.	0.376	0.402	5.170	28.32	22.01
Number of individuals	840	840	840	798	798

Notes: The sample is restricted to individuals in the immediate incentive treatment and individuals in the control group. Individuals in the cue/feedback treatment were not included in this analysis. All estimates include day of the week, week of the experiment, wave, and month fixed effects, baseline value of the outcome variable, indicators for the number of classes starting before 10am and demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Immediate Incentives and Sleep: Sensitivity Analyis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				sleep ≥ 7			sleep hours
	All	Has term GPA	No Missing	No wave 3	Pooled incentives	No Cue in Post	No Cue in Post
Treatment	0.1182***	0.1171***	0.1274***	0.1193***	0.1089***	0.1210***	0.3682***
	(0.013)	(0.013)	(0.012)	(0.014)	(0.012)	(0.021)	(0.058)
Post-Treatment	0.0556***	0.0554***	0.0554***	0.0639***	0.0530***	0.0636**	0.0903
	(0.015)	(0.015)	(0.014)	(0.016)	(0.014)	(0.025)	(0.059)
Immediate Cash*No Cue in Post-Treatment						0.0112	0.1034
						(0.027)	(0.064)
Observations	46,989	46,146	35,182	41,753	58,833	19,932	19,932
R-squared	0.256	0.254	0.158	0.259	0.254	0.263	0.302
Mean of Dep. Var.	0.429	0.427	0.432	0.434	0.435	0.455	6.742
Std. dev.	0.495	0.495	0.495	0.496	0.496	0.498	1.528
Number of individuals	840	825	840	763	1040	357	357

Notes: The sample is restricted to individuals in the immediate incentive treatment and individuals in the control group. Individuals in the cue/feedback treatment were not included in this analysis. All estimates include day of the week, week of the experiment, wave, and month fixed effects, baseline value of the outcome variable, indicators for the number of classes starting before 10am, and demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Immediate incentives and sleep: By quartiles of sleep at baseline

	(1)	(2)	(3)	(4)
Quartile Baseline Sleep ≥ 7	Q1	Q2	Q3	Q4
	Sleep$\geq$7			
Treatment	0.1493***	0.1099***	0.1513***	0.0976***
	(0.023)	(0.024)	(0.022)	(0.026)
Post-Treatment	0.0948***	0.0356	0.0408	0.0279
	(0.021)	(0.028)	(0.029)	(0.030)
Observations	11,999	11,907	8,763	8,729
R-squared	0.105	0.078	0.049	0.064
Mean of Dep. Var.	0.116	0.320	0.516	0.791
Std. dev.	0.320	0.467	0.500	0.407
Number of individuals	213	211	206	210
	Sleep hours			
Treatment	0.3571***	0.3484***	0.4126***	0.2754***
	(0.072)	(0.077)	(0.070)	(0.083)
Post-Treatment	0.2078***	0.1941**	0.2087**	0.0811
	(0.072)	(0.086)	(0.093)	(0.089)
Observations	11,999	8,906	8,763	8,729
R-squared	0.241	0.063	0.057	0.093
Mean of Dep. Var.	5.773	6.394	6.919	7.627
Std. dev.	1.378	1.444	1.466	1.150
Number of individuals	213	211	206	210

Notes: The sample is restricted to individuals in the immediate incentive treatment and individuals in the control group. Individuals in the cue/feedback treatment were not included in this analysis. All estimates include day of the week, week of the experiment, wave, and month fixed effects, baseline value of the outcome variable, indicators for the number of classes starting before 10am, and demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable in the control group at baseline. St.dev. is the standard deviation of the dependent variable in the control group at baseline. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Immediate incentives and Sleep: Heterogeneity

	(1) Male	(2) Female	(3) First-term	(4) Other students	(5) No-STEM major	(6) STEM major
Treatment	0.0941***	0.1303***	0.1493***	0.1116***	0.1150***	0 1249***
	(0.025)	(0.015)	(0.031)	(0.014)	(0.019)	(0.017)
Post-Treatment	0.0633**	0.0544***	0.1233***	0.0452***	0.0532**	0.0620***
	(0.026)	(0.018)	(0.036)	(0.017)	(0.024)	(0.020)
Observations	12,848	33,909	8,120	38,869	19,759	26,998
R-squared	0.229	0.256	0.258	0.257	0.256	0.255
Mean of Dep. Var.	0.344	0.463	0.424	0.430	0.476	0.397
Std. dev.	0.475	0.499	0.494	0.495	0.499	0.489
Number of individuals	229	607	160	680	355	481

Notes: All estimates include demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), baseline sleep, indicators for the number of classes starting before 10am, indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline.

Table A.7: Immediate incentives and Course Grade: Sensitivity Analysis

	(1) Baseline	(2) Basic controls	(3) Has grade in term+1 or term+2	(4) No missing HS/Baseline GPA	(5) Excludes wave 3 (Covid)	(6) Includes 8 obs with no baseline sleep	(7) Pooled incentives
Panel A: All classes							
Incentives	0.075**	0.064*	0.062*	0.067*	0.091**	0.075**	0.061*
	(0.037)	(0.038)	(0.036)	(0.037)	(0.040)	(0.037)	(0.035)
Observations	4,256	4,256	4,064	4,178	3,890	4,300	5,210
R-squared	0.129	0.103	0.131	0.134	0.130	0.128	0.126
Mean of Dep. Var.	3.499	3.499	3.498	3.500	3.488	3.502	3.495
Std. dev.	0.766	0.766	0.761	0.767	0.779	0.763	0.765
Number of individuals	825	825	784	808	749	833	1019
Panel B: Lectures							
Incentives	0.089**	0.076*	0.075*	0.079*	0.106**	0.088**	0.074*
	(0.042)	(0.043)	(0.040)	(0.042)	(0.045)	(0.042)	(0.040)
Observations	3,382	3,382	3,227	3,312	3,099	3,413	4,166
R-squared	0.137	0.108	0.139	0.143	0.138	0.136	0.134
Mean of Dep. Var.	3.434	3.434	3.433	3.434	3.421	3.436	3.433
Std. dev.	0.807	0.807	0.801	0.809	0.821	0.805	0.803
Number of individuals	819	819	780	802	744	827	1013
Panel C: Other Classes (Seminars, Labs etc.)							
Incentives	0.004	-0.006	0.001	0.003	0.014	0.004	-0.010
	(0.041)	(0.040)	(0.042)	(0.041)	(0.042)	(0.040)	(0.037)
Observations	874	874	837	866	791	887	1,044
R-squared	0.121	0.089	0.120	0.123	0.127	0.121	0.111
Mean of Dep. Var.	3.751	3.751	3.748	3.753	3.750	3.753	3.745
Std. dev.	0.507	0.507	0.513	0.505	0.505	0.505	0.519
Number of individuals	554	554	531	547	502	562	666

Notes: Estimates in columns 1, and 3-7 include demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), baseline sleep, indicators for the number of classes starting before 10am, indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). Basic controls only include gender, baseline sleep, and quartiles of high-school GPA or previous term GPA if high-school GPA was missing. For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline.

Table A.8: Immediate incentives and GPA: Heterogeneity

	(1) Male	(2) Female	(3) First-term	(4) Other students	(5) No-STEM major	(6) STEM major
			Panel A: All classes			
Immediate incentives	0.030	0.084*	0.163**	0.062	-0.013	0.132***
	(0.700)	(0.060)	(0.024)	(0.143)	(0.802)	(0.008)
Observations	1,125	3,110	772	3,484	1,754	2,481
R-squared	0.195	0.113	0.184	0.133	0.183	0.141
Mean of Dep. Var.	3.374	3.545	3.528	3.493	3.551	3.463
Std. dev.	0.856	0.725	0.705	0.778	0.722	0.793
Number of individuals	225	596	160	665	348	473
			Panel B: Lectures			
Immediate incentives	0.028	0.102**	0.194**	0.074	0.005	0.144**
	(0.745)	(0.045)	(0.025)	(0.120)	(0.927)	(0.011)
Observations	923	2,440	615	2,767	1,387	1,976
R-squared	0.214	0.121	0.229	0.138	0.195	0.157
Mean of Dep. Var.	3.322	3.477	3.465	3.427	3.496	3.391
Std. dev.	0.877	0.774	0.742	0.820	0.761	0.835
Number of individuals	223	592	160	659	344	471

Notes: All estimates include demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other), baseline sleep, indicators for the number of classes starting before 10am, indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline.

Table A.9: Immediate Incentives and Time Use (in minutes)

	(1) sleep	(2) sleep ≥ 7	(3) study	(4) social	(5) work	(6) eating & preparing	(7) exercise
Treatment	2.7484	0.0637***	-1.2500	-3.8814	-7.0458	-0.3915	1.2548
	(5.621)	(0.017)	(7.210)	(4.745)	(5.310)	(2.271)	(1.721)
Post-Treatment	9.8178	0.0570**	4.9436	-3.5697	-4.1311	-5.7144**	-3.3725
	(7.147)	(0.024)	(9.607)	(6.755)	(6.964)	(2.746)	(2.449)
Observations	5,970	5,970	5,970	5,970	5,970	5,970	5,970
R-squared	0.158	0.158	0.357	0.234	0.384	0.260	0.235
Mean of Dep. Var.	478	0.710	311.4	98.04	89.47	92.05	21.94
Std. dev.	138.4	0.454	197.5	122.6	142.8	53.09	42.36
Number of individuals	840	840	840	840	840	840	840
	(8) house errands	(9) personal care	(10) screen	(11) tv & other videos	(12) internet	(13) games	(14) other
Treatment	-0.7274	0.5681	-13.2889***	-8.7378***	-3.7657	-0.2943	3.6711
	(1.301)	(1.614)	(4.992)	(3.141)	(3.119)	(2.134)	(9.787)
Post-Treatment	-1.7599	-1.0092	-7.8137	-4.2680	-3.1771	-0.0613	-4.7491
	(1.693)	(2.325)	(6.965)	(4.441)	(4.217)	(3.336)	(12.296)
Observations	5,948	5,948	5,948	5,948	5,948	5,948	5,948
R-squared	0.115	0.279	0.402	0.318	0.403	0.350	0.187
Mean of Dep. Var.	17.52	52.33	166.4	68.59	76.39	21.40	486.6
Std. dev.	36.40	40.79	137.9	91.54	89.76	62.75	243
Number of individuals	836	836	836	836	836	836	836

Notes:
The sample is restricted to individuals in any of the cash incentive treatments and individuals in the control group. Individuals in the cue/feedback treatment were not included in this analysis. All the estimates include controls for day of the week fixed effects, month fixed effects, indicators for the number of classes starting before 10am, gender, race (dummies for Asian, Black, Hispanic, other) and ethnicity, parental education (dummies for less than college, college degree, and post-college degree), number of classes starting before 10am, quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing), and the average time spent on the activity at baseline. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline.

Table A.10: Immediate Incentives, Cognitive Performance and Physical health

	(1) Correct math answer	(2) Creativity score	(3) RHR	(4) # steps	(5) Active minutes
Treatment	0.0006	0.0011	-0.2564	12.9527	-1.6007
	(0.023)	(0.057)	(0.186)	(153.982)	(3.859)
Post-Treatment	-0.0359	-0.0086	-0.1761	-98.9606	-8.3426
	(0.032)	(0.060)	(0.205)	(235.088)	(6.085)
Observations	3,181	3,243	38,450	38,845	38,845
R-squared	0.187	0.045	0.836	0.359	0.296
Mean of Dep. Var.	0.363	3.307	65.73	8413	228.4
Std. dev.	0.481	0.717	8.245	5182	120.3
Number of individuals	809	803	832	840	840

Notes: The dependent variable in column 1 is an indicator equal to 1 if the respondent answered correctly the math question on the survey.The dependent variable in column 2 is a creativity score (see Section 2.3). RHR corresponds to participants Resting Heart Rate (RHR). Steps corresponds to participants' daily steps as measured via the Fitbit. Active minutes capture any activity at or above about 3 metabolic equivalents (METs). All estimates include day of the week, week of the experiment, wave, and month fixed effects, baseline value of the outcome variable, and demographic controls for gender, age (dummies), race and ethnicity (Asian, Black, Hispanic, White, other),indicators for the number of classes starting before 10am, indicators for whether parents' highest academic title was less than college, college degree, more than a college degree, and quartile of baseline GPA (high school GPA if non-missing, prior term GPA if high school GPA is missing). For all demographic characteristics, we included a missing indicator for whether the variable was missing. Standard errors are clustered at the individual level. Mean of Dep. Var is the mean of the dependent variable at baseline. St.dev. is the standard deviation of the dependent variable at baseline.. *** p<0.01, ** p<0.05, * p<0.1

Table A.11: Review of Post-secondary Interventions

Paper	Treatment	Setting	Findings	GPA locations
Angrist et al. (2009)	A) Prize money for academic achievement B) Peer advising and study groups C) Treatments A and B combined	Field experiment with first-year students at Canadian 4-year University	A) Insignificant GPA change (-.04, s.e. = .061) B) Insignificant GPA change (.011, s.e. = .063) C) Significant GPA increase (.168, s.e. = .086), decreased withdrawals (-.069, .036)	GPA: Table 6, Panel A, column 1 Costs: Bottom of page 160
Angrist et al. (2014)	Financial Incentives for academic achievement	Field experiment with students at public university in Ontario	Insignificant GPA change (.009, s.e. = .044)	GPA: Table 4b, "Fall" Panel, Column 9 Costs: Table 3, "Fall" Panel, Column 9
Barrow et al. (2014)	Receiving performance-based grant aid as well as counseling services as part of the Opening Doors Louisiana Program	Field experiment with low-income community college students in Louisiana	Significantly increases GPA (.182, s.e. = .085) and credits taken one semester later (1.234, s.e. = .30)	GPA: Table 8, Column 2 Costs: Table 2, "First semester" panel, Column 1
Clotfelter et al. (2018)	Receiving extra state-grant aid due to crossing income threshold for Carolina Covenant Grant eligibility	Regression discontinuity with low-Income students attending the University of North Carolina, Chapel Hill	Insignificant GPA impacts (.043, s.e. = .053), significant impacts on 4-year completion rates (.068, s.e. = .040)	GPA: Table 6, Panel A, Column 1 Costs: Table 3, Panel B, Column 1
Denning et al. (2019)	Receiving extra grant aid (Pell and State) due to crossing threshold for $0 Expected Familiy Contribution	Regression Discontinuity with 4-year University and Community College students in Texas	Insignificant GPA increases for FTIC (.031, s.e. = .026) and returning (.014, s.e. = .013) students, significant positive impacts on degree completion and later earnings	GPA: Table 3, Panel B, Column 3 Costs: Table 2, Column 2
Evans et al. (2020)	A) Access to emergency grant funding B) Treatment A as well as advising services	Field experiment with low-income community college students in Texas	A) Insignficant GPA changes (-.134, s.e. = .083) B) Insignficant GPA changes (.055, s.e. = .07), signficant increased enrollment amongst female students (.04, s.e. = .041)	GPA: Treatment A provided by authors, Treatment B from Table 8, Column 2 Costs: Pages 958-959

Table A.11: Review of Post-secondary Interventions (Continued)

Paper	Treatment	Setting	Findings	GPA locations
Goldrick-Rab et al. (2016)	Receiving extra need-based grant aid as part of the Wisconsin Scholars Grant	Field experiment with low-income first-year students at public universities in Wisconsin	Insignificant GPA changes (.08, s.e. = .06) and total credits completed (.9, s.e. = 1.7) for first cohort studied, and significant GPA increases (.09, s.e. = .03) and total credits completed (2.1, s.e. = .7) for cohorts 2 & 3	GPA: Table 5, "First Semester" Panel, Columns 2 & 5 Costs: Bottom of page 1772
Oreopoulos and Petronijevic (2018)	A) Online Exercise encouraging future-oriented thinking B) Treatment A as well as study advice and motivation via text messages C) Treatment A as well as one-on-one peer support	Field experiment with students at three campuses of the University of Toronto	A) Insignificantly changes 0-100 scale course grades (.143, s.e. = .575) B) Insignificantly changes course grades (.073, s.e. = .505) C) Significantly increases course grades (4.897, s.e. = 1.874) and credits completed (.501, s.e. = .283)	GPA: Table 3, Column 5[1] Costs: Bottom of page 323[2]
Park and Scott-Clayton (2018)	Receiving extra Pell grant aid due to crossing threshold for $0 Expected Familiy Contribution	Regression discontinuity with community college students from over 20 institutions in a single state	Insignificant GPA changes (.064, s.e. = .082), increases percent enrolled following year (.094, s.e. = .034)	GPA: Table 5, Column 2 Costs: Table, Column 2
Scott-Clayton (2011)	Receiving free tuition as part of the West Virginia PROMISE program	Regression discontinuity with public University students in West Virginia	Significantly increases GPA (.066, s.e. = .066), credits earned one year later (1.83, s.e. = .118), and 4-year completion rates (.067, s.e. = .005)	GPA: Table 3, Column 3 Costs: Middle of page 617

[1] Course grades on 100-point scale have been divided by 25 for comparison with 4.0 GPA scale
[2] Only treatment arm C is included in figure because costs could not be calculated for A and B

C. Instructions and Experimental Material

Immediate Incentives

PLEASE READ THROUGH THIS MESSAGE ENTIRELY
Starting this Sunday, and every weeknight (Sunday-Thursday) for the next five weeks, we encourage you to get 7 hours of sleep or more by 9am the following morning.
Every time you meet this goal (i.e., sleep 7 hours by 9am), you will earn a $4.75 PAYMENT via Venmo. Payments are redeemable only until 3pm on the days you earn them, and you will receive the payment by 3pm if you have redeemed by that time.

HOW IT WORKS
Every morning, you will receive feedback on your sleep. If you meet your goal, you will also receive the payment information via text message.

Next, we would like to ask you to pick your bedtime behavior – a behavior you would like to engage on right before going to sleep. Every weeknight, we will remind you of your bedtime behavior and we will encourage you to go to sleep early enough to meet your goal of sleeping at least 7 hours by 9am. Please pick your bedtime behavior by texting back the number of your choice. If you choose other, please type 9, then the behavior you want to set as your bedtime behavior.

1. Turn off your phone
2. Turn your phone to silent
3. Turn off your computer
4. Turn off Netflix
5. Turn on bedtime music
6. Turn on meditation app
7. Turn on white noise
8. Turn on pink noise
9. Other

Delayed Incentives

PLEASE READ THROUGH THIS MESSAGE ENTIRELY
Starting this Sunday, and every weeknight (Sunday-Thursday) for the next five weeks, we encourage you to get 7 hours of sleep or more by 9am the following morning.
Every time you meet this goal (i.e., sleep 7 hours by 9am), you will earn a $4.75 PAYMENT via Venmo. Payments are redeemable only until 3pm on the days you earn them, and the payment will be added to the amount of money you receive at THE END OF THE STUDY.

HOW IT WORKS
Every morning, you will receive feedback on your sleep. If you meet your goal, you will also receive the payment information via text message.

Next, we would like to ask you to pick your bedtime behavior – a behavior you would like to engage on right before going to sleep. Every weeknight, we will remind you of your bedtime behavior and we will encourage you to go to sleep early enough to meet your goal of sleeping at least 7 hours by 9am. Please pick your bedtime behavior by texting back the number of your choice. If you choose other, please type 9, then the behavior you want to set as your bedtime behavior.

1. Turn off your phone
2. Turn your phone to silent
3. Turn off your computer
4. Turn off Netflix
5. Turn on bedtime music
6. Turn on meditation app
7. Turn on white noise
8. Turn on pink noise
9. Other

Cue / Feedback

PLEASE READ THROUGH THIS MESSAGE ENTIRELY
Starting this Sunday, and every weeknight (Sunday-Thursday) for the next five weeks, we encourage you to get 7 hours of sleep or more by 9am the following morning.

HOW IT WORKS
Every morning, you will receive feedback on whetehr you met your goal.

Next, we would like to ask you to pick your bedtime behavior – a behavior you would like to engage on right before going to sleep. Every weeknight, we will remind you of your bedtime behavior and we will encourage you to go to sleep early enough to meet your goal of sleeping at least 7 hours by 9am. Please pick your bedtime behavior by texting back the number of your choice. If you choose other, please type 9, then the behavior you want to set as your bedtime behavior.

1. Turn off your phone
2. Turn your phone to silent
3. Turn off your computer
4. Turn off Netflix
5. Turn on bedtime music
6. Turn on meditation app
7. Turn on white noise
8. Turn on pink noise
9. Other

Creativity Instructions (Example)

You will be asked to complete different short tasks over the course of the study. One of these tasks will be chosen for payment at the end of the study.

Today's task: Using some or all of the words below, write an interesting sentence. Your sentence will be rated based on its creativity from 1-5 points, where 5 is the most creative. If today's task is chosen for payment, your payment will be determined by how creative your sentence is. **You will receive $1 for each point your story is rated**.You will receive as little as $1 for completing this activity and up to $5 for the most creative sentences. You will receive your rating and your payment at the end of the study.

The words for you to use in your sentence are:
(Example) event, chocolate, system, indicate, article, emotion, possess, mom, poetry, reality

Math Instructions (Example)

You will be asked to complete different short tasks over the course of the study. One of these tasks will be chosen for payment at the end of the study.

Today's task: On the next page you will be asked to answer a math question. If today's task is chosen for payment, your payment will be determined by whether you answer the question correctly, and how quickly you answer. **You will receive $1 for answering the question correctly, and you will receive an additional $0-$4 depending on how quickly you answer the question**. You will receive as little as $1 for answering this question correctly and up to $5 for the quickest correct answers. You will receive your score and the payment at the end of the study.

Here is the question you are asked to answer:

It costs a manufacturer X dollars per component to make the first 1,000 components. All subsequent components cost $1 each. When X = $1.50 How much will it cost to manufacture 4,000 components?

o $3,500
o $3,000
o $4,000
o $3,250
o $4,500

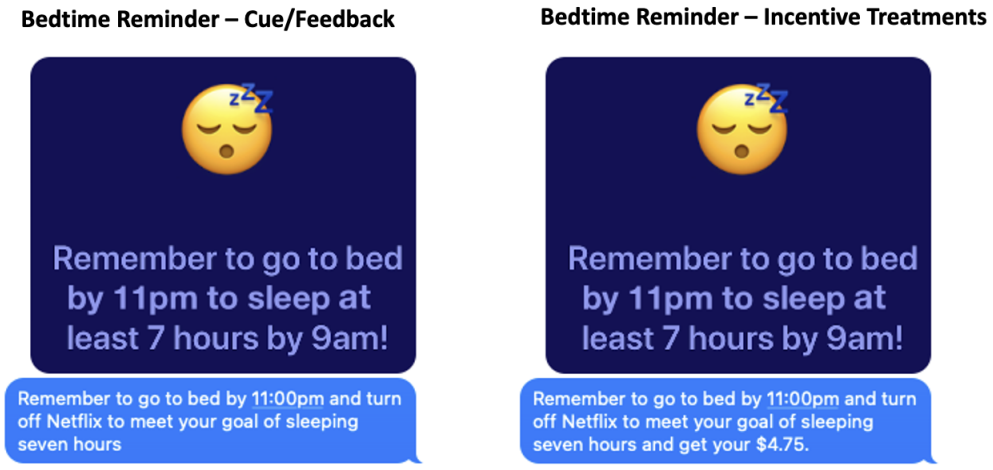
Cost-Effectiveness Analyisis

To ensure comparability with our results, GPAs reported in Figure 6 are non-cumulative, either at the semester or year-level. All coefficients included in the figure are OLS estimates of program impacts. Whenever applicable, we use authors' baseline estimate for program

impacts. To calculate per subject per semester costs, we divide total program costs by the ITT sample size of the program. See Table A.11 for more information on each study included.

App ScreenShots

Figure C.1: Bedtime Reminder



Notes - The Bedtime reminder included a personalized goal bedtime of approximately 1 hour before the baseline bedtime, with a latest possible time of 1am. It also included a personalized bedtime behavior participants chose from before the beginning of the intervention, from a list containing "Turn off your phone", "Turn your phone to silent", "Turn off your computer", "Turn off Netflix", "Turn on bedtime music", "Turn on meditation app", "Turn on white noise", "Turn on pink noise", "Other". If participants selected "Other"m they could specify a behavior of their choice.

Figure C.2: App Screenshots - Immediate Incentive Treatment

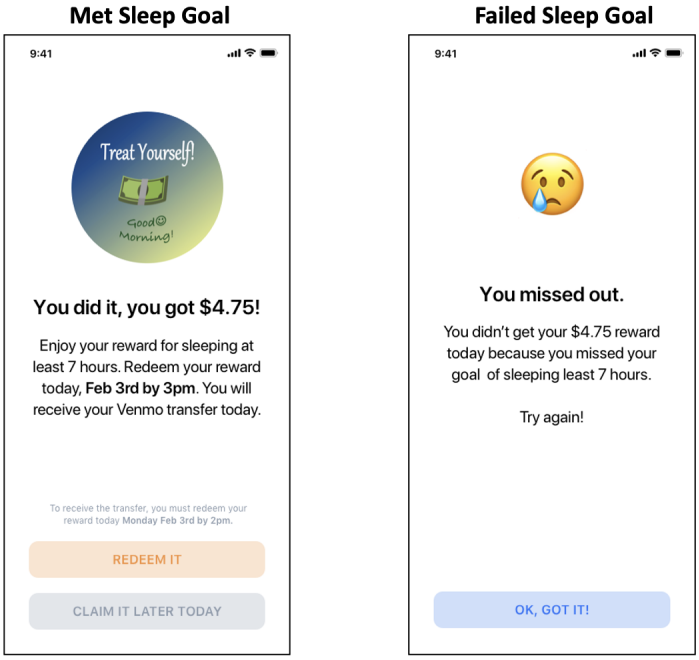


Figure C.3: App Screenshots - Delayed Incentive Treatment

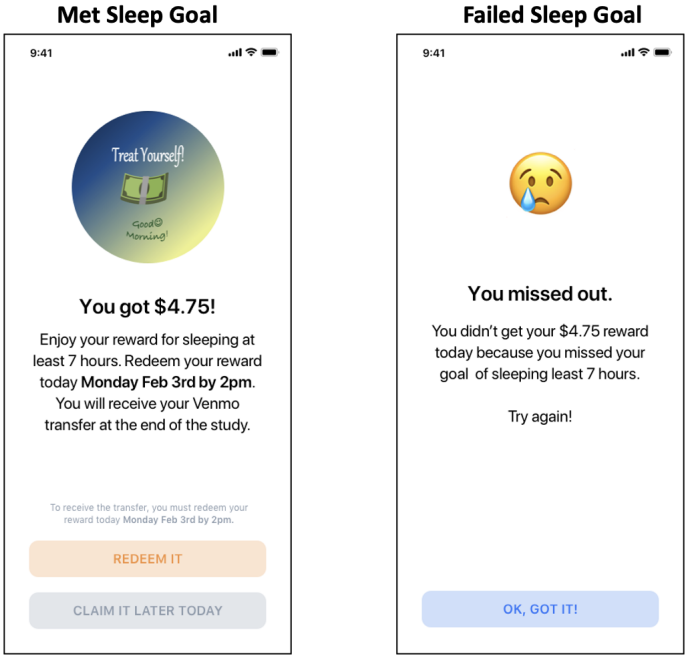


Figure C.4: App Screenshots - Cue/Feedback Treatment

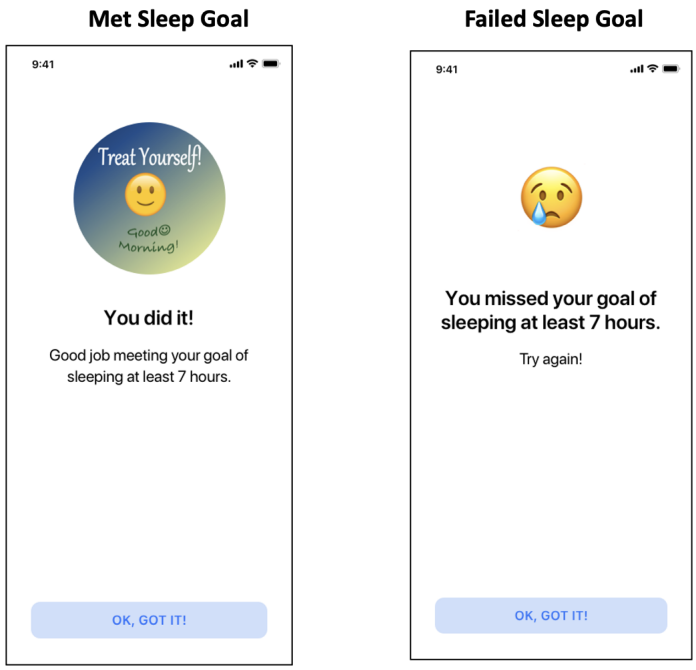


Figure C.5: Reminder to Sync - All Treatments

