Sleep Study Experiment

W241 Final Project

David Hou
Ravi Ayyappan
Ahsen Qamar

20 April 2019

Introduction

The importance of good sleeping habits has been emphasized frequently in contemporary advice. One aspect of good sleep hygiene is the avoidance of any electronic devices with backlit screens prior to bed. Light from electronic devices can cause the brain to stay active as if it is a daytime thereby inhibiting the production of the sleep hormone 'melatonin'. In this experiment, we seek a quantitative measure of sleep quality degradation as a result of screen-time before bed. Due to our limited expertise and resources for studying sleep science, we will perform the measurement using a mobile app.

Research Question

Does using electronic devices with a backlit screen thirty minutes prior to sleep have a statistically significant effect on the sleep quality as measured by a mobile app?

Hypothesis

We hypothesize that that use of an electronic device with a screen for at least thirty minutes prior to going to sleep degrades the overall sleep quality for that night. The null hypothesis would correspondingly say that there is no statistical significance between screen time thirty minutes before going to bed and sleep quality.

As a correlary, we investigate whether refraining from electronics usage thirty minutes prior to going to sleep improves sleep quality for that night.

Experimental Design

The mobile app that will be used for the experiment is called Sleep Cycle and it utilizes the phone's microphone to measure the user's sleep cycles while in bed. It is primarily advertised for its alarm clock feature, which attempts to wake a user in the lightest phase of sleep within a preset time interval. Each morning, the app displays a graph of the user's sleep cycles for the previous night and some relevant statistics. Figure 1 shows an example of the output from one night's data.
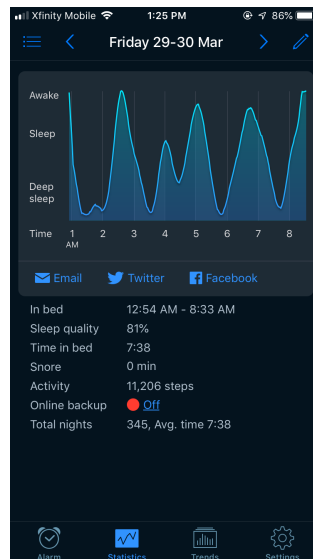

Figure 1: Example night

One of the statistics is the "Sleep Quality" number, which is the principle outcome variable for this experiment. According to the *Sleep Cycle* website, the following four factors go into calculating Sleep Quality:

1. Amount of time spent in bed.
2. Amount of time spent in deep sleep.
3. Consistency of the sleep.
4. Amount of times where the app registered the person as fully awake.

Of the four factors, users generally have direct control only over the first one. We did not control the amount of time that users spend in bed, but we did instruct subjects to try to sleep at regular hours every night. As will be discussed later, compliance is a major issue in an experiment of this sort. We felt that tightly controlling the amount of time in bed would exacerbate the problem. The other three factors are hopefully affected by screen-time, allowing us to measure a treatment effect. The app also records a "snore" and "steps walked" value, but we have doubts about the accuracy of these measures and will not use them in our analysis.

The experiment took place over two weeks, from March 23 to April 5 of 2019. Saturday was

chosen as the start day of the week because it was feared that instructions to change the sleeping habits in the middle of the week might be ignored. We recruited 35 people for the study using a Google survey (provided in the appendix), consisting of classmates, friends, and family. In addition to asking for contact and demographic information, the survey also inquired about general sleep habits such as bedtime and pre-sleep activities. While the survey did ask about electronic usage, the questions also included workout habits and alcohol consumption. Prior to commencement of the experiment, subjects were not made aware of the fact that screen-time was the critical variable of interest.

All communications to subjects occurred over email. Every person in the same experimental group received identical emails (see Appendix). When instructing subjects to use electronic devices prior to bed, we did not mandate any specific activity, as we felt it would increase non-compliance on an already-intrusive study. We only specified that users should use some sort of electronic device with a backlit screen for 30 minutes immediately before bedtime. Similarly, we did not ask subjects in the other group to perform anything in lieu of electronic usage. These subjects were simply asked to refrain from using electronics with backlit screens 30 minutes before bedtime.

Subjects were randomly assigned to two groups, with group 1 receiving treatment (screen-time before bed) during the first week and group 2 receiving control (no screen-time before bed). After the first week, subjects were instructed to swap to the opposite activity of their original assignment. This swapping design allows us to perform within-subject comparisons while also leaving time for treatment effects to manifest themselves. With the group that received treatment first, we are also able to perform some analysis on persistence effects. Note that subjects were not informed of the instruction swap ahead of time, so we do not expect any anticipation effects.

The nature of the experiment requires effort from subjects in both treatment and control groups. That is, for people who normally use electronic devices before bed, the request to refrain from usage constitutes a significant change in pre-sleep ritual; for people who normally avoid electronics before bed, the opposite is true. Therefore, we are concerned with two-sided non-compliance, which is especially difficult to measure in this experiment because people are unlikely to honestly report when they have deviated from our instructions. To help combat this to an extent, we offered subjects some small monetary compensation and emphasized prioritization of honesty while reporting compliance data. Even so, we can never be sure of whether subjects followed our instructions or not.

Experiment Deviations

After the initial sign-up survey, we asked all respondents to begin using the app (from March 14 to March 22 in a normal way so that we can collect the baseline data. Further, the app requires a minimum usage of 5 days to calibrate the sleep quality number. Sleep habits during this baseline usage were not closely controlled or monitored, as each subject signed up for the experiment on different dates. However, we did ask for an export of this baseline data prior to starting the experiment, so that subjects had some awareness of the steps it takes

to perform the export. At this time, we learned that Android users could not perform the export due to missing functionality in the app. As such, we asked Android users (4 out of the 38) to manually record the sleep times and quality from the app each day.

After closing out the initially sign-up survey, we decided to include ourselves as well as subjects in the experiment. This was to slightly increase the number of subjects. Though we were not aware of our own randomization apriori, we obviously had more knowledge of the experiment than other subjects, so we will analyze our own data with special care.

Results

Data Cleaning

We made as little modifications to the raw data as possible, but a few exceptions exist. We decided to omit entries for which the subject slept for a duration shorter than one hour. On a case by case basis, we determined that these entries were caused by improper usage of the app. There were several entries where the user had two entries for the same night separated by some period of awakeness. For these cases, one or both the periods of sleep were also very short. However, none of them fell under the one hour threshold for omission.

Of the 35 people who originally signed up for the study, only 29 people (including 3 of us) actually submitted the final data by the time of writing this report. The remaining 9 people are not necessarily attritioners, as we could not collect data from them from the very beginning (no baseline or treatment or control data available from them).

Analysis

Experiment Power

Before presenting the analysis, it should be noted that this experiment is very underpowered. It was difficult to recruit a significant number of people who would agree to have their sleep monitored and controlled over several weeks. For our `n` size of 29, to observe a small effect size (Cohen's `d = 0.2`) at the 0.05 significance level, we can only expect to have a power of about 0.12.

Sleep Quality

Figure 2 shows the distribution of the sleep quality over the entire experiment. This includes data from the baseline and actual experiment. We see a slight left skew, even after removing the entries with fewer than one hour of sleep. There is also a slight build up at 100 due to that being the maximum number that the app reports. We do not believe that the peaks around 70 and 80 are noteworthy.
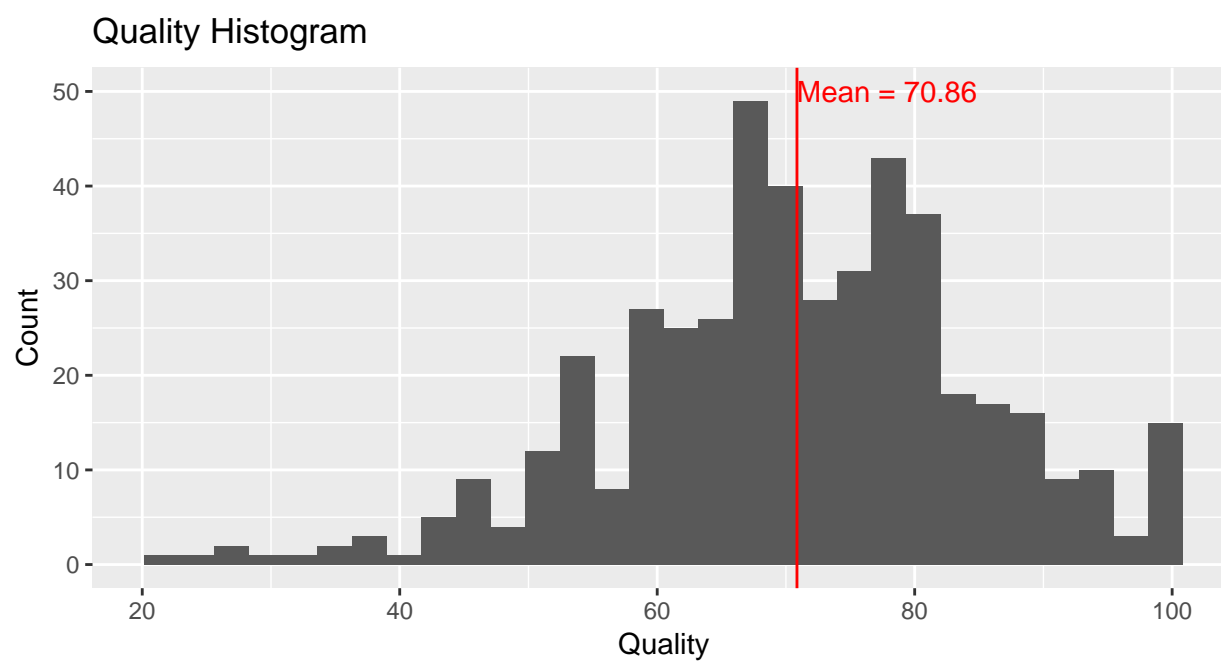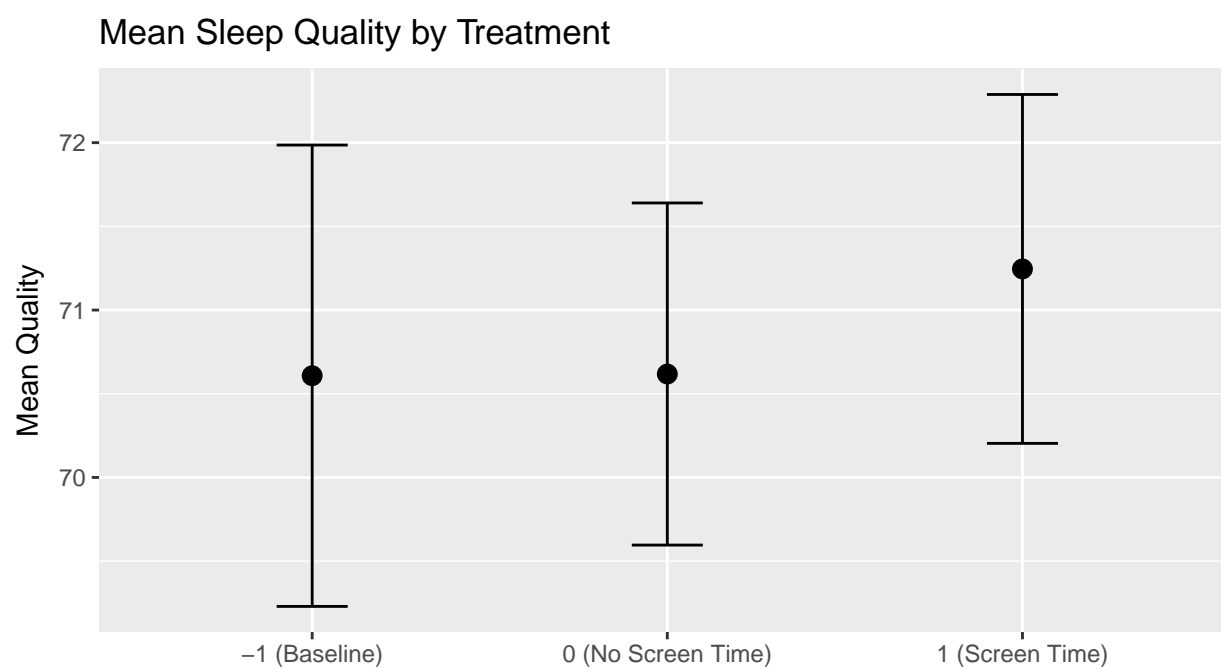
Quality Histogram

Figure 2:   Quality histogram

Mean Sleep Quality by Treatment

Figure 3:   Mean sleep quality by treatment

Table 1: Mean sleep quality by treatment

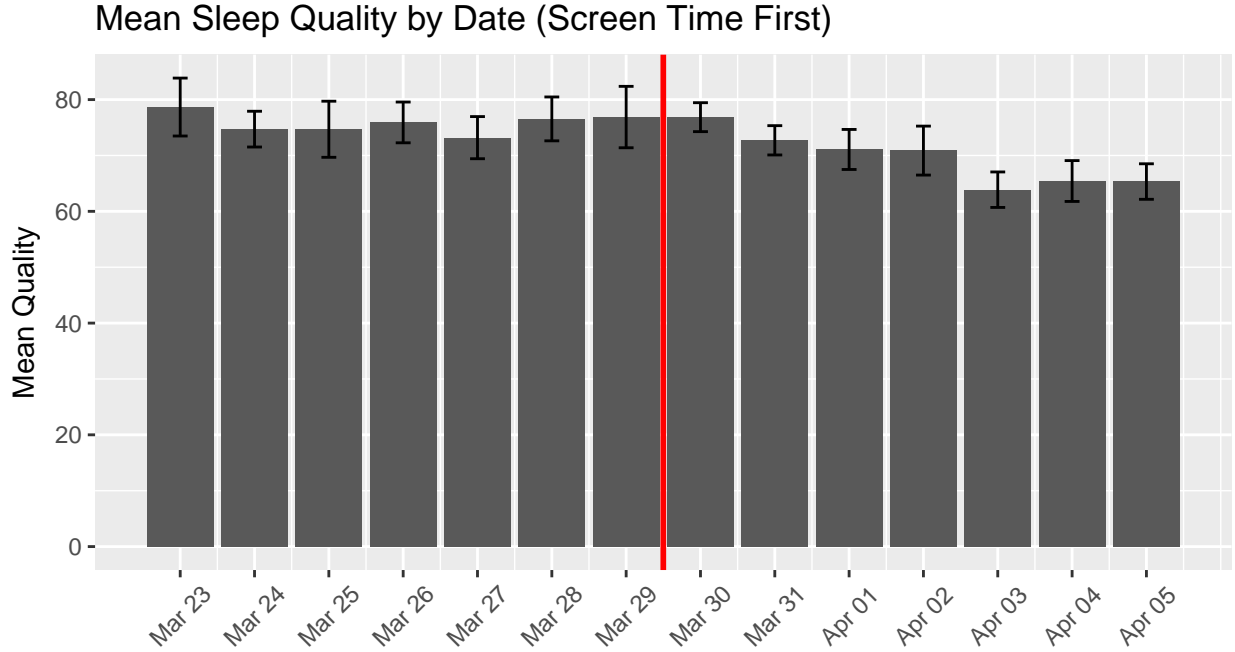| Label	Mean	St. Err.	N
Baseline	70.608	1.378	130
No Screen Time	70.618	1.022	157
Screen Time	71.246	1.042	179

Figure 4: Mean sleep quality by date in group which received screen time first

Figure 3 shows the sleep quality, averaged over the respective time periods, for each treatment assignment. The reported error bars are standard errors for each time period. As mentioned above, the baseline data was collected in the time between subjects signing up and the experiment officially starting. Recall that that we did not tightly control baseline data and not every subject provided it.

It is already clear that the experiment will not show the hypothesized effect, as the point estimate for screen time is actually slightly higher than that with no screen time. More importantly, the standard errors completely dominate any difference between the treatments. Table 1 shows the tabular results for Figure 3.

Figure 4 and Figure 5 show the mean sleep quality by dates for the two groups. Recall that group 1 received treatment (screen time) in week 1 of the experiment and switched to control (no screen time) in week 2. Group 2 had the opposite schedule. Again, the large error bars make it hard to draw any sort of conclusion for whether the switch in treatment has any effect, but it could be argued that there is a slight drop in quality in group 1, after they switch from screen time to no screen time. Although still likely due to random noise, this drop is probably what primarily drives the treatment point estimate being in the opposite
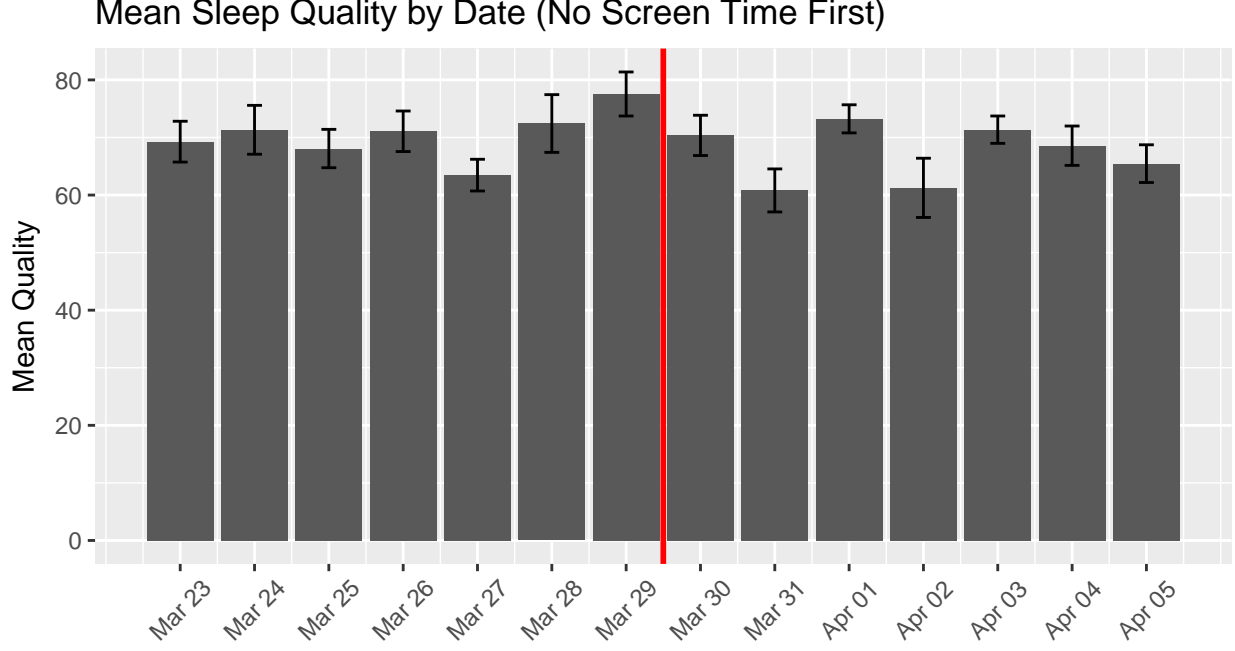
Figure 5: Mean sleep quality by date in group that received no screen time first

direction of the expected effect.

Table 2 shows the results of three regressions models:

$$quality = \beta_0 + \beta_1 treat$$

$$quality = \beta_0 + \beta_1 treat + \sum_i \gamma_i id_i$$

$$quality = \beta_0 + \beta_1 treat + \sum_j \delta_j x_j$$

In model 1, we naively regress on just the treatment variable. This yields the 0.628 point effect that was already shown in Table 1. However, we are not taking into the account that multiple entries in the dataset came from the same person.

We assigned a numerical ID to each participant in the study. This provides a unique indicator variable for each person and the standard errors reported in Table 2 are actually clustered standard errors based on those IDs. In model 2, we control for the fixed effects to each person. This brings down the treatment effect (which was already statistically insignificant) even closer to zero.

Recall that the sign-up form for the experiment also contained several questions regarding sleep habits. In model 3, we regress on these in lieu of the subject IDs. See the pre-experiment survey in the appendix for the list of covariates. Interestingly, the treatment effect now flips to be the hypothesized sign. We don't emphasize this flip any further, as we are still far from

7

statistical significance.

The self-reported bedtime stands out as a variable that is has a significant coefficient. We will not draw any causal inference from this fact, as we did not randomize subjects' bedtimes. It is also important to reiterate that these are not the actual observed bed times in the experiment, but the estimated bed time that subjects reported in the pre-experiment survey. However, it does seem that for this group of people, self-reported late sleepers also have higher sleep quality.

The other significant coefficient is the one associated with sharing a bed. We are not surprised by this finding. Unless both people are simultaneously using the app, *Sleep Cycle* cannot distinguish movement from one person or another. It should be the case that multiple people should make much more noise and trick the app into thinking that its user is awake for longer than actuality.

Table 2: Regressions, outcome is sleep quality

	Dependent variable:		
	Sleep Quality		
	(1)	(2)	(3)
Treatment	0.628 (1.581)	0.120 (1.639)	−0.464 (1.606)
Age			0.288 (0.199)
Male			1.555 (3.027)
Bed time			0.401*** (0.141)
Rise time			0.669 (1.666)
Daily caffeine			−1.417 (1.113)
Workout			−0.695 (2.484)
Daily screen			−0.033 (0.317)
King size bed			4.624* (2.526)
Queen size bed			−4.911** (2.383)
Twin size bed			4.126 (3.680)
Shares bed			−9.470*** (3.427)
Constant	70.618*** (1.564)	71.940*** (0.819)	58.471*** (19.735)
Control for user ID	No	Yes	No
Observations	336	336	326
R^2	0.001	0.297	0.145
Adjusted R^2	−0.002	0.231	0.112
Residual Std. Error	13.424 (df = 334)	11.761 (df = 306)	12.554 (df = 313)
F Statistic	0.183 (df = 1; 334)	4.460*** (df = 29; 306)	4.424*** (df = 12; 313)

Note: *p<0.1; **p<0.05; ***p<0.01
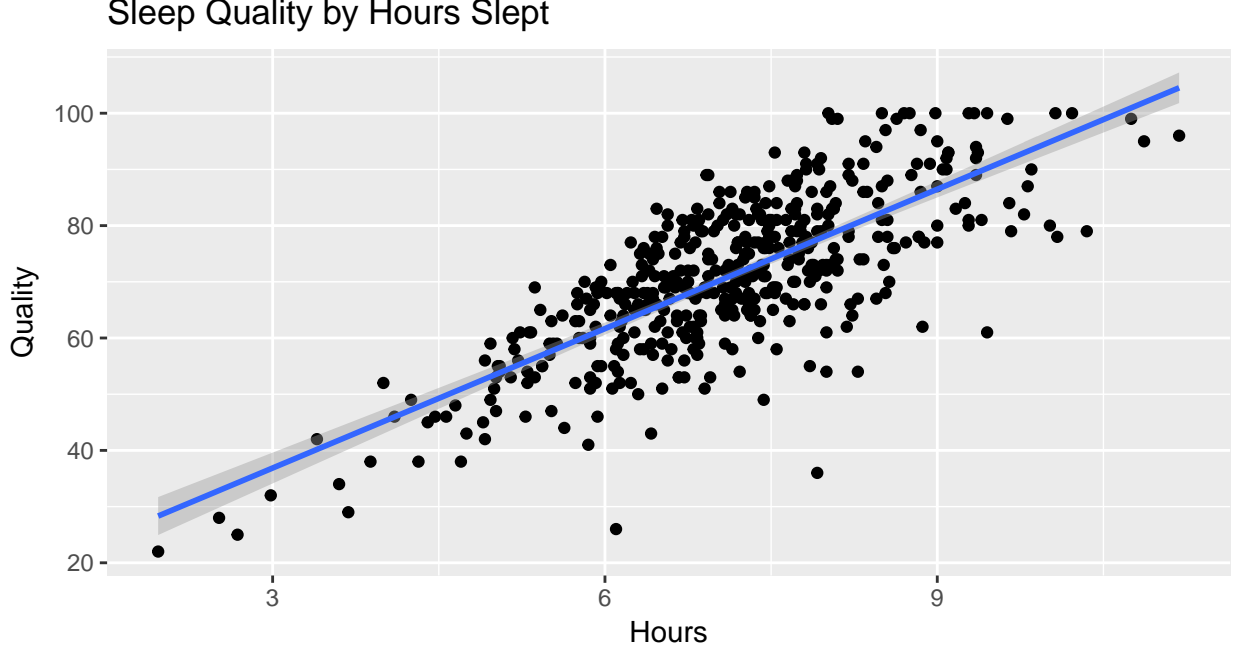
Figure 6: Mean sleep quality by hours slept

Hours Slept

Figure 6 shows that sleep quality is highly correlated with hours slept. Table 3 shows the strength of this relationship. Since these two variables are so predictive of one-another, it may be interesting to also use hours slept as an outcome variable. Note that we did not include hours slept as a covariate in the previous section because it was measured post-treatment and would be a bad control if included.

Thus, we ran the analogous three regression to Table 2:

$$hours = \beta_0 + \beta_1 treat$$

$$hours = \beta_0 + \beta_1 treat + \sum_i \gamma_i id_i$$

$$hours = \beta_0 + \beta_1 treat + \sum_j \delta_j x_j$$

Table 4 shows the results. Again, we see no statistically significant treatment effects, but find that several more of the covariates are now significant. Note that treatment seems to increase hours slept, regardless of other controls. This is likely the reason we are unable to find the hypothesized effect in sleep quality. Ideally, we would have been able to tightly control the amount of time each subject spent in bed.

Since the coefficient on estimated daily bed time is still significant, we investigate the effect

Table 3: Regression, quality on hours

	Dependent variable:
	quality
Hours	8.268***
	(0.396)
Constant	12.060***
	(2.853)
Observations	466
R^2	0.584
Adjusted R^2	0.583
Residual Std. Error	9.088 (df = 464)
F Statistic	650.944*** (df = 1; 464)
Note:	*p<0.1; **p<0.05; ***p<0.01

of treatment on actual observed bed times.

Bed Time

Figure 7 shows a finding that maybe interesting. During baseline collection, it appears that the subjects slept slightly later than during the experiment. Since we obviously cannot observe other people who did not participate in the experiment, we refrain from drawing any causal conclusions. In any case, it is apparent that the actual treatment had little effect on bed time.

Experimenter's Effect

Finally, we present the effect of being an experimenter in the results. Recall that we included ourselves in this study. Table 5 shows that knowing about the details of the study did not seem to have any significant effect on the outcome variables.

Non-compliance

Non-compliance was a major issue in this experiment. Beyond not following our instructions on electronics usage, many subjects simply did not use the app. Figure 8 shows the number of nights of data that each subject actually submitted. The experiment took place over two weeks and ideally we would have received 14 nights of data from each person. Clearly this was not the case. The mean number of nights submitted was just 11.6 days, with one participant

Table 4: Regressions, outcome is hours slept

	Dependent variable:		
	Hours Slept		
	(1)	(2)	(3)
Treatment	0.123 (0.126)	0.088 (0.138)	0.024 (0.145)
Age			0.047** (0.021)
Male			0.275 (0.312)
Bed time			0.042** (0.018)
Rise time			0.212 (0.193)
Daily caffeine			−0.058 (0.131)
Workout			−0.327 (0.227)
Daily screen			0.0003 (0.033)
King size bed			0.386 (0.352)
Queen size bed			−0.588* (0.317)
Twin size bed			0.959** (0.433)
Shares bed			−0.994** (0.406)
Constant	7.054*** (0.157)	7.272*** (0.069)	4.065* (2.195)
Control for user ID	No	Yes	No
Observations	336	336	326
R^2	0.002	0.398	0.176
Adjusted R^2	−0.001	0.341	0.144
Residual Std. Error	1.243 (df = 334)	1.009 (df = 306)	1.149 (df = 313)
F Statistic	0.819 (df = 1; 334)	6.988*** (df = 29; 306)	5.555*** (df = 12; 313)

Note: *p<0.1; **p<0.05; ***p<0.01

Bed Time by Treatment
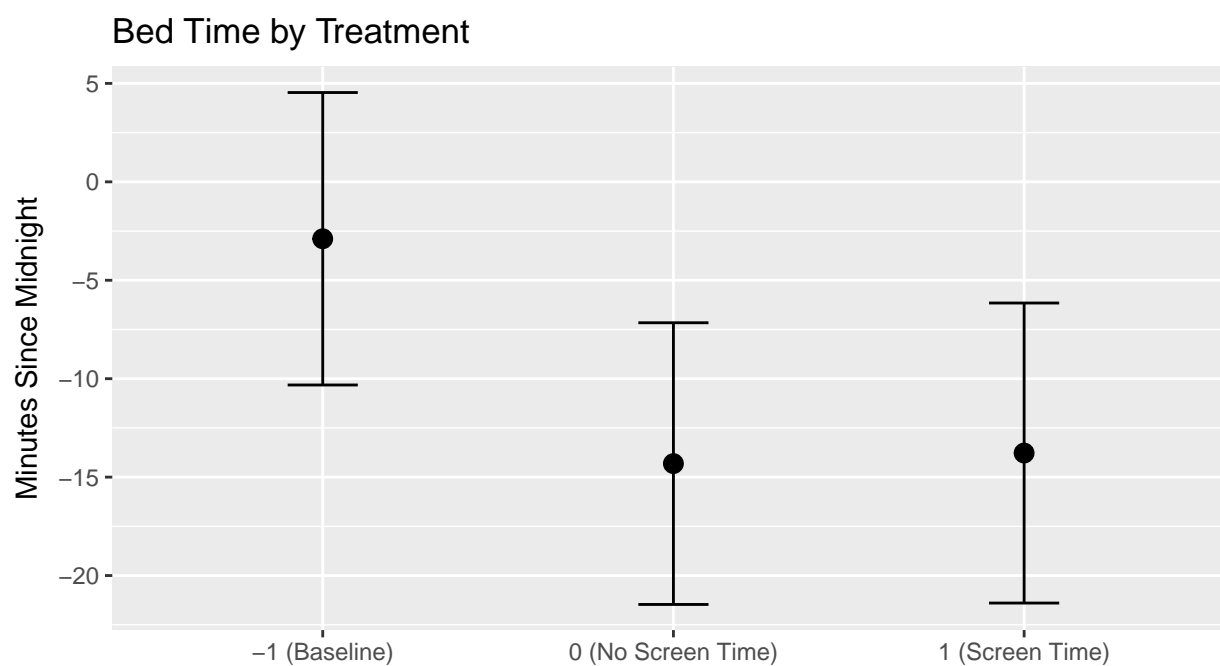
Figure 7:   Bed time by treatment

Data submitted by each subject
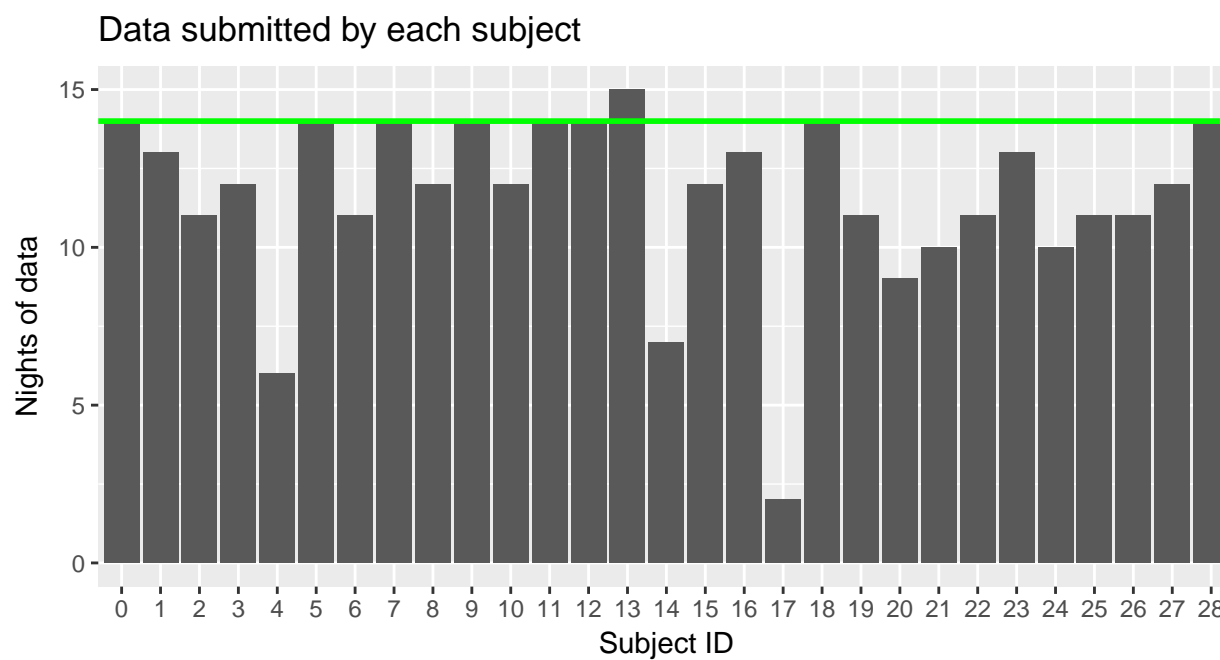
Figure 8:   Number of nights of data submitted by each participant

Table 5:   Regressions, effect of being an experimenter

	Dependent variable:	
	Sleep quality	Hours slept
	(1)	(2)
Treatment	−0.121	0.140
	(1.659)	(0.133)
Experimenter	−0.759	0.391
	(3.672)	(0.652)
Treatment*Experimenter	6.730*	−0.162
	(3.734)	(0.349)
Constant	70.700***	7.012***
	(1.713)	(0.158)
Observations	336	336
R^2	0.011	0.009
Adjusted R^2	0.002	−0.0002
Residual Std. Error (df = 332)	13.393	1.243
F Statistic (df = 3; 332)	1.255	0.974

Note: *p<0.1; **p<0.05; ***p<0.01

only using the app for 2 out of the 14 nights as he had a baby ahead of time. As an aside, one subject with 15 nights of data used the app twice on one night. The subject went to sleep, woke up, and then went to sleep again.

Challenges

Throughout the course of this experiment we encountered challenges both expected and unexpected. We had hoped that incentivizing participants with a $10 gift card would help recruitment but it did not. Right before starting the experiment we had some participants dropout resulting in an even smaller group for our experiment. It would be a good idea to invest some time in understanding the best way to recruit commited participants.

Another fact that we were made painfully aware of is that people do what they want and they do not like being told what to do. The figure below shows self-reported anonymized non-compliance from the participants:
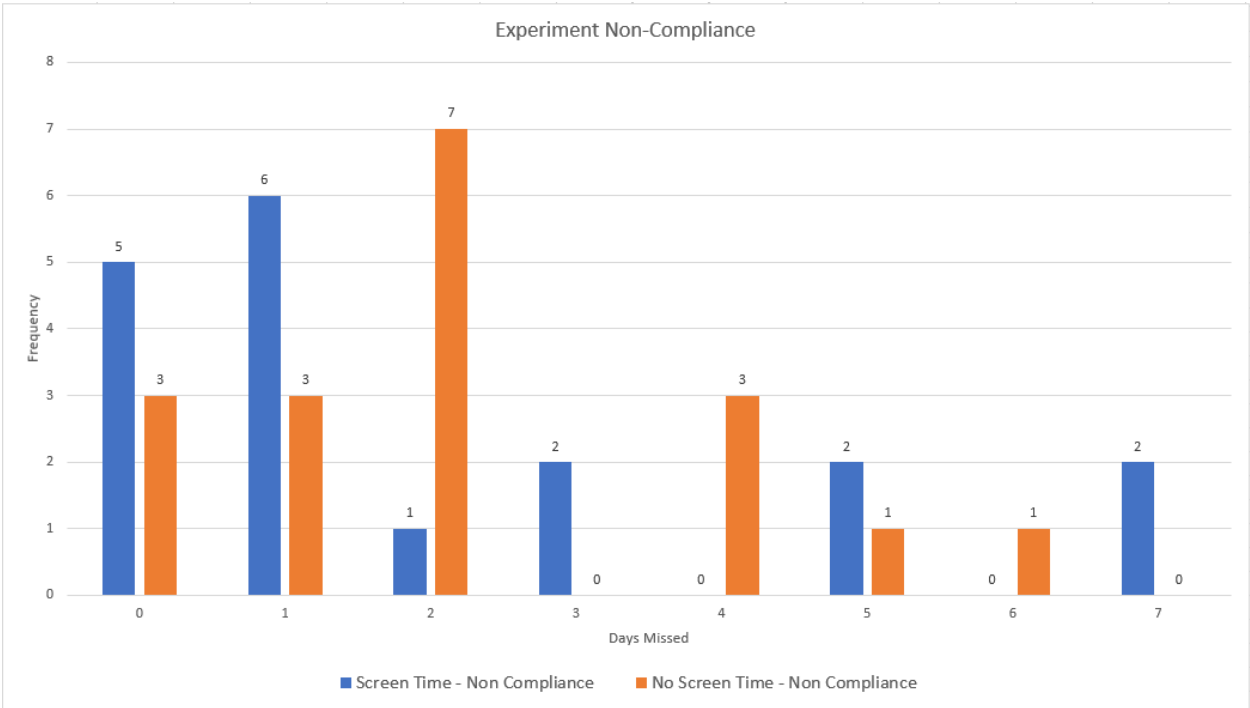
Figure 9: Experiment Non-compliance

We see from Figure 9 that we had two participants that were not able to follow the treatment for all seven days. Majority of the participants missed around one to two days. Reflecting on these numbers from an experiment point of view leads us to few ways the experiment could have been conducted better to reduce non-compliance. We could have sent nightly reminders to participants reminding them to follow treatment and control. However, how would this apply across participants who may have different bedtimes. We could have filtered users

14

with abnormal sleep cycles (i.e. medical residents) out of the survey. If this experiment was repeated special attention needs to be paid to whom we are recruiting.

Towards the end of the experiment questions also came up regarding our decision to use the *Sleep Cycle* app. When it came to analyze the data we saw that the app reports higher sleep quality numbers proportional to time spent in bed. Additionally, the Android version of the app was lacking the ability to export data and had to be reported manually from the participants. The *Sleep Cycle* app uses audio analysis to discern and generate a sleep quality number. There is no way to no way to know for sure if this is the best method to measure sleep quality. Another alternative is to use the product created by *Withings*: https://www.withings.com/us/en/sleep#.

There is a strong possibility that our experiment suffered from the spillover effect. A decent amount of our participants were recruited from word of mouth or relatives of those conducting the experiment. No explicit instructions were sent to individuals to not share whether they were assigned to treatment (screen time) or control (no screen time). There is no way to know for sure but a bigger pool of participants and more explicit instructions would greatly reduce the risk.

Conclusion

Through our findings and analysis of the data we fail to reject the null hypothesis. This experiment was likely too small to demonstrate any statistically significant treatment effect. It is also interesting to note that the treatment increased the number of hours slept, even when controlling for other covariates. This study would be much more interesting if we could have tightly controlled the amount of time each subject spent in bed and we were able to recruit more subjects.

Appendix

This section contains the pre/post survey questions and the email correspondences that we had with the participants.

Initial Survey

Figure 10: Initial Survey

How many minutes prior to bed do you usually stop using any electronic devices?

○ 3

○ 15 min

○ 30 min

○ 45 min

○ 60 min

○ More than 90 min

Do you watch movies/videos or play video games before going to bed?

○ Movies/Videos

○ Play video games

○ Both

○ None

Do you browse news sites or social networking sites before going to bed?

○ Yes

○ No

○ Maybe

What bed size do you have?

○ Twin

○ Queen

○ King

○ California King

Figure 11: Initial Survey

Welcome Email

Welcome to the sleep study!

w241 sleepstudy <w241sleepstudy@gmail.com>   Fri, Mar 15, 10:00 PM

Hello!

Welcome to the sleep study and thank you so much for participating! We are still gathering the last few participants but would like you to at least download the app and get used to using it. If you haven't already, install the app from https://www.sleepcycle.com/ or the App Store or Google Play or whatever fancy place you get apps from. You'll find that the interface is pretty intuitive, but more information on how to use the app can be found here https://www.sleepcycle.com/how-sleep-cycle-works/. We'd like to collect as much data as possible, so if you could start using the app as soon as possible, that would be awesome. Don't worry that it's a weekend and you don't want an alarm. The app has a setting where it will just track your sleep quality but not wake you on weekends.

Thanks again for agreeing to participate and we'll inform you of more instructions when we want you to try something specific.

Happy dreams!

The sleep study team

Figure 12: Welcome Email

Week 1 - Control group Email

Experiment starts from today. Please follow the below instructions !!!

Inbox ×

w241 sleepstudy <w241sleepstudy@gmail.com>   Sat, Mar 23, 4:16 PM

Hello!

Good evening.

Hopefully you have downloaded the app and have been using it for a while. Please email your baseline sleep data from Sleep cycle app -> Settings -> Advanced -> Database -> Export Database to w241sleepstudy@gmail.com. This will send the 'sleepdata.csv' as an attachment using the mail client configured on your phone.

For this upcoming week - *Mar 23 to Mar 29 (till next Friday)*, we would like you to continue using the app as usual (by setting the alarm and dismissing it in the morning) every single night to track the sleep quality. You can go to bed and wake-up as per your normal routine. *The only caveat is that we would like you to refrain from using any devices with backlit screens 30 minutes prior to sleep.*

For some of you, we know this could be very difficult, but please try your best to comply with these instructions. We understand that we are all human beings and impulse control is tough at times. If you do end up using any electronic devices (however long they were) within 30 minutes of going to bed, please make a note of the date and let us know. This study won't last forever and we really appreciate your time and effort!

The sleep study team

Figure 13: Week1 Control Email

18

Week 1 - Treatment group Email

Experiment starts from today. Please follow the below instructions !!! Inbox ×

w241 sleepstudy <w241sleepstudy@gmail.com>   Sat, Mar 23, 4:23 PM

Hello!

Good evening.

Hopefully you have downloaded the app and have been using it for a while. Please email your baseline sleep data from Sleep cycle app -> Settings -> Advanced -> Database -> Export Database to **w241sleepstudy@gmail.com**. This will send the 'sleepdata.csv' as an attachment using the mail client configured on your phone.

For this upcoming week - **_Mar 23 to Mar 29 (till next Friday)_**, we would like you to continue using the app as usual (by setting the alarm and dismissing it in the morning) every single night to track the sleep quality. You can go to bed and wake-up as per your normal routine. **_The only caveat is that we would like you to use some electronic device for 30 minutes before going to bed._** This could include anything from watching shows on your laptop/TV/tablet device, checking social media on your phone, playing video games etc. Really anything with a backlit screen ON.

For some of you with good sleeping habits, this would be a departure from your normal ritual and we apologize for the deviation. We know this could be very difficult, but please try your best to comply with these instructions. If you don't end up using any electronics device 30 minutes before going to bed, we would like you to make a note of the date and let us know. This study won't last forever and we really appreciate your time and effort!

The sleep study team

Figure 14: Week1 Treatment Emaill

Week 2 - Control group Email

Sleep Study: Updated instructions - Last week of the experiment Inbox ×

w241 sleepstudy <w241sleepstudy@gmail.com>   Sat, Mar 30, 2:34 PM

Hello!

Good afternoon.

Thanks for continuing to use the app. This will be our last week of the experiment.

Please email your sleep data so far.
- iPhone users: Sleep cycle app -> Settings -> Advanced -> Database -> Export Database to **w241sleepstudy@gmail.com**. This will send the 'sleepdata.csv' as an attachment using the mail client configured on your phone.
- Android users: Please provide the "Date","Sleep quality number","Total time in bed" in a spreadsheet and mail it to us at **w241sleepstudy@gmail.com**.

For this upcoming week -**_Mar 30 to Apr 5 (till next Friday),_** we would like you to continue using the app as usual (by setting the alarm and dismissing it in the morning) every single night to track the sleep quality. You can go to bed and wake-up as per your normal routine. **_The only caveat is that we would like you to refrain from using any devices with backlit screens 30 minutes prior to sleep._**

For some of you, we know this could be very difficult, but please try your best to comply with these instructions. We understand that we are all human beings and impulse control is tough at times. If you do end up using any electronic devices (however long they were) within 30 minutes of going to bed, please make a note of the date and let us know.

We really appreciate your time and effort!

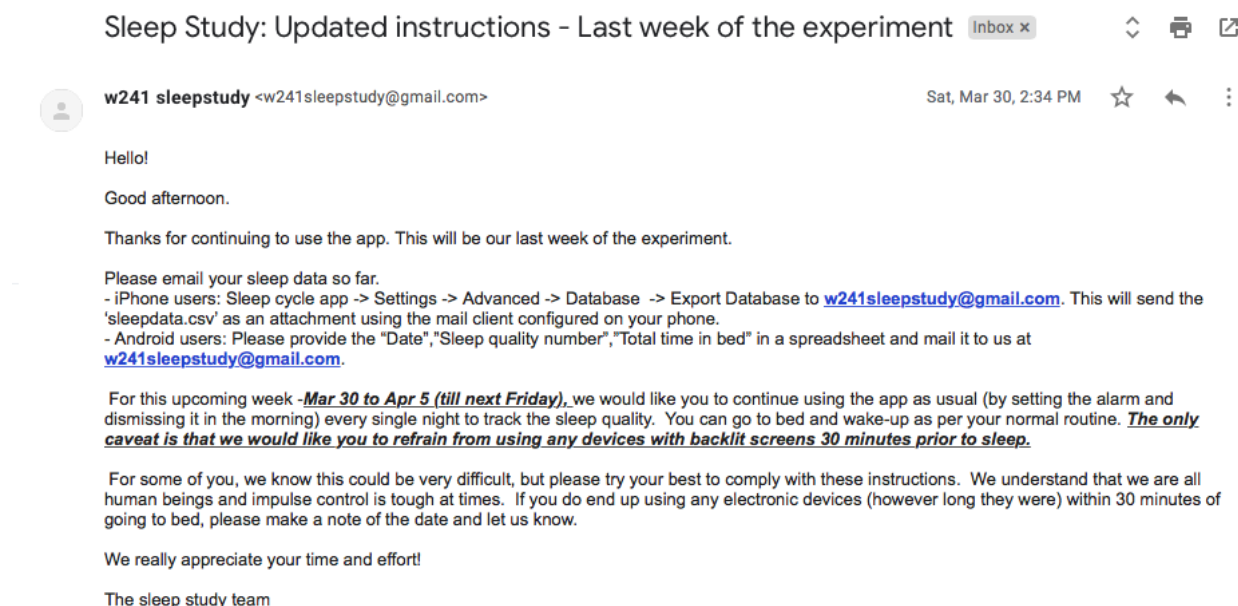The sleep study team

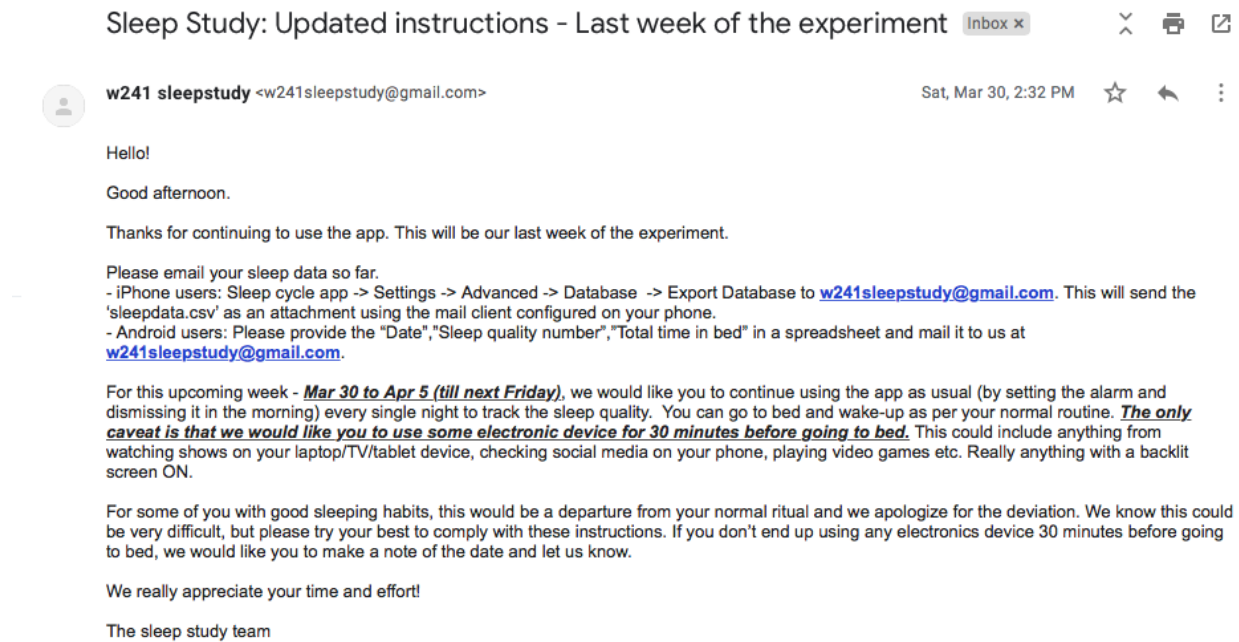Figure 15: Week2 Control Email

Week 2 - Treatment group Email

Figure 16: Week2 Treatment Email

Post Experiment Survey

Sleep Study Follow Up

Hello! This is a very simple, two question form where we need some information for you as participants in the study. Please note that responses are completely anonymous. We cannot tell who is responding to this at all so please give as honest answers as possible! Thanks again for participating!

On days you were asked to spend time using a device with a screen 30 *

- ○ 0
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7

Figure 17: Post Survey questions

On days you were asked NOT to spend time using a device with a screen 30 *

- ○ 0
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7

Figure 18: Post Survey questions