Sleep Health

An Android App Platform For Improving Sleep Health With BCSS and UX design

An Interactive Qualifying Project submitted to the faculty of

Worcester Polytechnic Institute

In partial fulfillment of the requirements for the degree of Bachelor of Science

by:

Alonso Martinez (Robotics & Computer Science),
Akshay Thejaswi (Computer Science),
Michael Perrone (Mathematics),

December 17, 2015
Submitted to:
WPI Advisor: Professor Vance Wilson
Co-Advisor: Professor Soussan Djamasbi
Sponsor: N/A

Contents

List of Figures

List of Tables

Abstract

In this project, the concepts of Behavior Change Support Systems (BCSS) and responsive design were reviewed, and these theories/methodologies were applied to the development of a phone app with the goal of helping to improve sleep health amongst a college undergraduate population. The phone app provides reminders and useful information, allows personalization of the features and interface, praises and empowers the user when they exhibit positive behavior, and provides feedback and goal tracking features. These aspects of the app were used to try and project expertise while still promoting similarity, providing praise and rewards systems, facilitating self-monitoring, and providing reminders as they are defined within the context of BCSS. The design and testing of the app was documented so that subsequent student teams may expand upon the research and the Android app that serves as the foundation for testing the robustness of BCSS and addressing the problem of poor sleep health among WPI undergraduates.

1 Executive Summary

At the beginning of the project, a goal was set to create an app that used various features to assist the user in improving their sleep patterns and health. The concepts underlying behavior change support systems were analyzed, and various features were conceptualized, drawing inspiration from BCSS.These features were then discussed and worked out using traditional software practices. Development of the features was done in different iterations that lasted a week. This practice of breaking up the work followed the Agile Development Cycle that many companies and professionals use when developing a product. The programming done also drew from User Experience design concepts. The interface and data visualization were programmed with minimalism in mind in order to ensure that the application was simple to use. Thus, by combining the concepts of BCSS and UXDM, an initial app was developed that was simple to use, and visually appealing.

A system of user-set alarms was implemented with customizable music. This enables the user to set reminders for themselves and personalize and customize the app to their liking. A system for self-reporting bedtimes and wake-up times with a button press was implemented as a minimally invasive way to gauge the phase and duration of sleep during the day. This feature can also be used to turn alarms off in the morning, or play soothing music as the user goes to sleep at night, providing utility to the user. Additionally, getting the user into the habit of clicking the button before bed will also help create a moment of self-reflection as the user is made to consciously notice and gauge their sleep decisions. Sleep quality surveys were implemented, both to allow the user to reflect on their sleep patterns in the last week, and to collect additional data. A database was implemented where all the collected data is stored for analysis. Analysis of this data was not implemented, but the database was created with analysis of smaller-scale databases in mind. Analysis can, in the future, help provide better feedback to the app users, assess the sleep situation amongst app users, and potentially discover new sleep health solutions. Graphing of the collected data is available for the user to look at, to help them keep track of and analyze their own sleep behaviors. These systems can be used to help support behavioral change in students

After the features and interface were programmed, two rounds of usability testing were done to get feedback about how to improve the app. The first round of testing brought to light many issues, allowing the developers to pinpoint problems in the code and remedy them. This resulted in another iteration of the app that had many of the initial problems fixed. While the Android platform is easy enough to understand and program for, there are many differences in the phones that run Android and how they run the app. Future teams should be wary of platform compatibility issues in the future, although all such existing problems have been more or less fixed. With the initial bugs out of the way, a second round of user testing was done in order to discover any remaining bugs. The results from the second round of testing were very positive and very few problems were reported. Another iteration of the app was created in order to fix the final remaining issues. The testing that was done was very important to the project, as the final version of the app would serve as a foundation for the future teams, so the app needed to be fully functional and debugged in order to ensure that tests could be run at a later time.

2 Acknowledgements

We would like to thank the professors in charge of the project for their support and enthusiasm for the project:

- Professor Soussan Djamasbi for her guidance and expertise with interface design, for writing advice, and for continued support and enthusiasm
- Professor Vance Wilson for his guidance with implementation and structure of the Android app, project advice, and proofreading skill

3 Introduction

3.1 Poor Sleep Health Among Students

The application was developed based on BCSS theory with the pressing matter of poor sleep health in mind. *Adolescent Sleep Needs and Patterns* goes into great detail about the problems and risks associated with poor sleep health in adolescents [4]. According to the report, driving while drowsy increases the risk of a car crash, and other injuries due to lack of alertness. The article also explains that poor sleep health may also lead to poor grades and performance in school, reduced short term memory and learning ability, negative moods, a higher probability of stimulant abuse, and a loss of some forms of behavioral control [4]. It is noted in the report that sleep is especially important in puberty, and daytime sleepiness increases and undergoes a phase delay, with teens going to bed later and waking up later [4]. Amongst college students, poor sleep health really is a problem: Kathryn Orzech's research shows that poor sleep health is a real problem for students residing on college campuses, and that it can negatively affect their health, wellbeing, and academic performance. The PSQI Score that was used had an average consistently around 6.3 while the standard deviation was usually around 2.7 [14]. A score of 5 or higher usually indicates a clinically diagnosable sleep disorder, and the results indicate that perhaps two thirds of the college students surveyed were in that range of 5 or higher. Thus, there exists a moral and practical imperative to help college students stay healthy and do their best in our society: it helps improve their quality of life, and it will strengthen the economy in the long run, because college age students are the future of the labor force.

There are methods by which healthy sleeping conditions can be promoted. *Adolescent Sleep Needs and Patterns* makes a number of recommendations about how schools, families, and communities can create supportive systems to help adolescents maintain healthy sleep behaviors. These include modifying the start time of the school district, educating students about the benefits of sleep and the risks of a lack of sleep, and keeping the home environment quiet late at night [4].

In this paper, focus is placed specifically on software interfaces as a structure for improving sleep health, because software solutions for health and wellness based on interaction between humans and computers are gaining academic interest fast and show great promise for improving people's lives [16]. There are a number of examples of software interfaces in the literature that can affect the sleep behaviors, or general health and wellness of the software users. In Lu's paper, it can seen that narrative health blogs invite the user to empathize with the characters in the story and learn more effectively from them [10]. Other research shows an improvement in sleep behaviors via a software system of persuasive reminders and virtual rehearsal [8]. An additional paper describes an effective software system for enabling self-monitoring of sleep habits with a very high compliance rate and low data capture burden[5]. Thus, for this project, a software solution that will serve as a foundation will hopefully have a great positive impact on the sleep health and wellness of students at WPI, and eventually students in colleges everywhere.

4 Background

The areas of study most informative to design decisions of the initial app were the young field of Behavior Change Support Systems, and the more mainstream User Experience Design. The iterative testing of the Android app used industry standard usability testing techniques. Here these areas of study are explained in greater detail.

4.1 BCSS and Persuasive Systems Design

Persuasive Systems Design is the broad concept that computerized systems can be used to persuade and influence the behavior of people. The concept is nuanced but broad, and it is discussed in detail in Fogg's book, *Persuasive Technology: Using Computers to Change What We Think and Do* [6]. There are many ways to implement persuasive systems, but common among them is a logical or emotional appeal or message built into how the computer's user interface is structured, which is intended to influence the user's decisions and behaviors.

Behavioral Change Support Systems are a particular set of persuasive system designs which try to enable the software user to make behavioral changes by supplying them with the tools and framework they need to achieve a behavioral change more easily. Behavioral Change Support Systems are collectively a relatively young offshoot of persuasive systems design, but there is much theory to support their efficacy in principle [12]. In practice, it is necessary to further test the robustness of such systems in real world applications. If they can perform as well as other persuasive systems, then they may be adopted more widely as behavior changing mechanisms. The app designed in this paper is intended as a platform for testing BCSS in real-world applications.

Persuasive Systems Design terminology was used to describe the types of Behavior Change Support Systems that were employed: "Personalization" describes two types of systems. In the more widely-used sense, interfaces may collect data about a user from interaction with that user, and analyze that data to try and adapt their form to suit the user's needs. This sort of automatic personalization is used by many modern, well-known websites like Facebook or Amazon. The other sort of personalization is that which allows the user to modify the form of the interface to what they feel suits their needs [13]. With the databases developed in this project, the first form of "Personalization" could eventually be implemented. The second form can describe the settings and alarms features implemented in the app developed during this project. "Expertise" describes persuasive systems that provide useful knowledge and experience. The utility of provided services and the applicability of provided information are what support and facilitate behavioral change [13]. For example, providing advice to the user based on analysis of the data they input could be considered a BCSS that uses expertise. Systems that provide "Reminders" aim to keep their users alert about their behavior when using the system, compared to how their goal behavior should be [13]. System-generated notifications when certain data patterns are observed can be considered a reminders-based BCSS. Systems which utilize "Praise" congratulate the user for good performance and use positive reinforcement to encourage more of the desired behaviors [13]. "Self-monitoring" systems provide the user with a means to gauge their own behavior [13]. Choe et al. implemented self-monitoring systems for users to gauge their sleep health [5], and in this project, the graph display feature allows users to track their sleep health. "Rehearsal" BCSS provide tools which enable users to act out or rehearse a target behavior, as implemented by Salam et al. [8] where the effect of virtual rehearsal on improvement of sleep behaviors was explored. Persuasive systems that use "Similarity" try to make the user empathize and learn from the situation of someone who is similar to them in some way [13]. Similarity effects can come from narratives, as during Lu's research on the persuasive effects of narratives in health blogs [10], but they can also come from real social situations, which is why it would be interesting to add features almost like social networking to a user interface, like how Fitbit users compete for high scores.

4.2 User Experience Design

User Experience-oriented Media Design is a field of study which analyzes the way people interact with interfaces on computers and other devices. Marc Hassenzahl discusses design oriented around the user experience at length in his influential book, "The Thing and I" [7]. In the third chapter of this book, Hassenzahl explains how developers experience the app in terms of features and implementation specifics, but users care about appeal, pleasure and satisfaction derived from the use of the app as well as objective utility. In practice this means that users will use interfaces that enable them to be productive, or they will use interfaces that they enjoy, or some reasonable mix of the two. The design decisions for the initial app were informed by this idea of the user experience. It is understood, for example, that a potential app user will only make use of the app if the value derived from its use is greater than the time and effort needed to use it. When making design decisions, it was also clear that a minimally complicated interface with a pleasant appearance was necessary for a good user experience.

4.3 Usability Testing

There are some well understood User Experience design principles for interfaces between computers and people, but often the actual user experience and interface effectiveness is difficult to predict in its entirety. In industry it is considered good practice to make sure user interfaces perform as expected. There are a number of performance metrics for user interfaces [11], but for this project, Task Success Matrices, expectation measurement, and the System Usability Scale were deployed in order to measure overall effectiveness. [1, 15].

The idea of a Task Success Matrix is to split the app interface up into tasks which are the individual interactions the user can perform within the app. A record is then kept of how many people successfully complete each task. This allows identification of which tasks in the app are most difficult to use, since they will have the lowest task success rates. These features can then be focused on when making improvements to the interface [15].

To measure the user expectation about the difficulty of an interface [1], the task the app is designed to accomplish was described, and then the focus group participants were asked to rate how difficult they think that app would be to use on a scale of 1 to 10. The participants are then asked to try the app themselves, and once they are done with that, they are asked how difficult to use the app actually was on the same scale of 1 to 10. This allowed the developers to gauge each member's subjective feelings about task complexity compared to the expected task complexity.

The System Usability Scale is an industry standard way of quantitatively gauging the usability of an interface based on user agreement with various positive and negative statements about interface usability. Study participants were allowed to try the interface out themselves, then they are given the system usability scale in a standard format to fill out. More info can be found in appendix C. SUS is thought to be quite reliable, and has an average score of 70.91 among about 200 studies using SUS since 1996 [9, 2, 3].

Usability testing is part of the formative design approach, wherein an initial structure can be designed, then iteratively tested and improved. The above usability testing methods were used to gain feedback to guide improvements and further design decisions.

5 Methodology

An Android app informed by User Experience design concepts, structured to help test the robustness of Behavior Change Support Systems and help improve sleep health among WPI students was developed. A formative design approach was taken, building an initial app structure, then iteratively testing it in focus groups, as usability studies to help debug and refine the app.

5.1 Implemented App Features

The decisions that influenced how the interface of the app utilized UX concepts and Behavior Change Support systems is described. The technical implementation and code structure is discussed the technical documentation. As an overview for future teams, the appearance of an Android app is specified by XML files, while the back end is usually Java code, with the whole setup being reminiscent of web development. Android app development is well documented, with plenty of resources available via a quick Google search. The code that was developed was stored in Git repositories on Bitbucket in a private repository, and Android Studio was used to program and debug the app, because it has a number of helpful features for Android app development. Initially the app had to be downloaded from a website, but this caused compatibility issues because various Android phones had non-default file downloading utilities, so for the second usability test the app was hosted on the Android app store instead.

When starting the app for the first time, the user is prompted to enter some initial information, like their age, the sleep goals they want to achieve, and an initial run through the Epworth Sleep Scale test (see appendix D). This data serves as an initial benchmark of sleep health and is stored for future use. Informed by UX design concepts, the amount of input is limited in the setup to the most vital information, allowing the user to voluntarily input data at their own pace later. This was done in order to lower the barriers to entry for potential app users. Social networking sites like Facebook and Linkedin use a similar strategy to get users to fill out their profile. The app then continues to collect information about sleep behavior in various ways that (again informed by UX design concepts) try to be simple, minimally invasive, and when possible, useful to the user. After the initial setup the user can also take the Epworth Sleep Scale test again weekly to track their sleep health over time. For gathering further information, the app has a button the user clicks when they wake up and go to bed, to track sleep duration and variability over time. This button also starts a tune to help lull the user to sleep, or to help wake the user up in the morning. The app stores that information in a database, so that it can be retrieved as CSV files and analyzed, both within the app to provide the user with feedback, and outside the app to enable researchers to test the robustness of Behavior Change Support Systems.

Getting into the habit of pressing that button also helps the user notice better when they are going to bed or waking up, stimulating a self-monitoring system to support behavioral change. There are other behavioral change support systems implemented in the app as well. The app allows the user to view the graph of their data over time, compared to their sleep goals, allowing more long-term self-monitoring and self-reflection. The app also allows the user to set reminders and alarms, to serve as a system of rehearsal and keep the user thinking about their goals. In principle all information collected can be analyzed to help the app decide what messages to provide the user. A default set of messages is implemented now, but an expanded form of this functionality will be a versatile BCSS, with messages to the user chosen by analysis of their sleep health data over time. In a general sense, messages chosen by data analysis can provide the user with a form of automated personalization to give the user the most relevant feedback, analogous to how Amazon chooses its list of products the user might be interested in based on data the website collects about the user. For example, when an improvement in sleep health is detected, the app can provide praise to support those healthy changes in sleep behavior. If the user is not doing so well in terms of sleep health, the app can provide helpful advice and extra reminders to the user, and provide relevant papers and resources to educate the user about the benefits of improved sleep behavior and promote an appearance of expertise to persuade the user to take the app advice seriously. For example, if one entered information saying that they drink a lot of coffee at night, the app can provide literature on the effects of sleeping caused by caffeine consumption and prompt the user to set reminders not to drink coffee after a certain time of night. To 'personalize' the app in terms of customization, the user can opt in or out of various app features (which doubles as an easy way to set up absence-presence testing), and the user can opt out of messages from the app for certain times of day, so that messages do not disturb the user at important times of day.

5.2 Focus Group Testing

The Android app was improved iteratively by testing via two focus groups, spaced a few weeks apart with some time in between to incorporate feedback into the app. Some demographic questions were collected, and users were asked for initial expectations vs. subsequent perceptions of difficulty level to gauge the difference between expectations and reality [1]. Tasks were then given to the user to complete. Notes were collected and recorded in a task completion matrix to get a sense of which parts of the app had more problems [15]. This data was then used to target refinements to the app. Open-ended feedback and thoughts were encouraged in questions after the tasks were completed. These features were compiled into a script and given to the participants in the focus group. This script can be found in appendix A. The participants were also given an industry standard System Usability Survey (see [3, 15] and appendix C) to help evaluate the usability of the app relative to other, similar user interfaces. In the second focus group, additional information regarding the type of Android phone the participants used was documented, in order to help track any compatibility issues in the Android platform. In addition, some minor modifications to the second wording of the focus group script were made: the new version can be found in appendix B.

6 Results

In the first usability test that was organized a focus group with four Android phone users who are students at Worcester Polytechnic Institute, two of whom were 20, and the remaining two were 21 and 22 years old. The group consisted of 3 women and 1 man, one in Environmental Engineering, another in Civil Engineering, and two in Management and Information Systems. All the students considered themselves experienced Android phone users. This was a small sample size to start with, but this was meant to be a formative study, with many of the issues being obvious enough at this point in development that detailed statistical analysis was not necessary to see what needed fixing. Sometimes with focus groups, one person takes charge of the particular issues that the group focuses on, but this was not observed in the focus group, with each participant exploring with some autonomy from the group, but seeking help when they were confused or ran into an issue.

A number of usability metrics were recorded in the first usability test. The expectation measure described by Albert and Dixon [1] was used, where the app objectives were described. The users were then asked to rate how difficult they thought the app would be based on their impression of the app, on a scale of 1 to 10, with 1 being very easy, and 10 being very difficult. Then, after letting users try the app, they were asked to rate the difficulty level they experienced. Table 1 shows the expected difficulty and the actual difficulty encountered, as reported by the users:

Table 1: Expected vs. actual difficulty levels in first study

Difficulty	participant 1	participant 2	participant 3	participant 4
Expected	4	8	7	3
Actual	8	9	5	4

During the study, participants were given a number of tasks to complete within the app, During this process, it was recorded whether they completed the process (with a 1), whether they were partially successful (recorded with a 1/2), or whether they could not accomplish the goal (recorded as 0). This data was recorded as a modified task success matrix [15] in table 2:

Table 2: Success rates of tasks in first study

Subject	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
1	0	1/2	0	0	1	1	1	1
2	1	1/2	1	0	1	1	1/2	1
3	1	1/2	1	1/2	1	1	1	1
4	0	1/2	1/2	0	1	1/2	1	1

Where the tasks in table 2 are drawn from the focus group script in appendix A in the table are as follows:

Figure 1: Focus group 1 tasks

1. Download and install the app off of arc30.wpi.edu/sleephealth
2. Proceed through the initial setup proc of the app
3. Set a time for an alarm
4. Set music for an alarm
5. Set a target sleep goal
6. View the graph of sleep data
7. Take the Epworth Sleep Scale Survey
8. Opt out of notifications/alarms

In addition, the industry-standard System Usability Survey [15] was used to evaluate the overall usability of the app by the time of the first focus group. The survey form used was as described by John Brooke [3](see appendix C). SUS Scores were calculated by subtracting one from the result of the odd questions and summing the result, then subtracting the even answers from 5 and summing that result, then multiplying by 2.5.

Figure 2: Graph of System Usability Survey negative prompts for first study

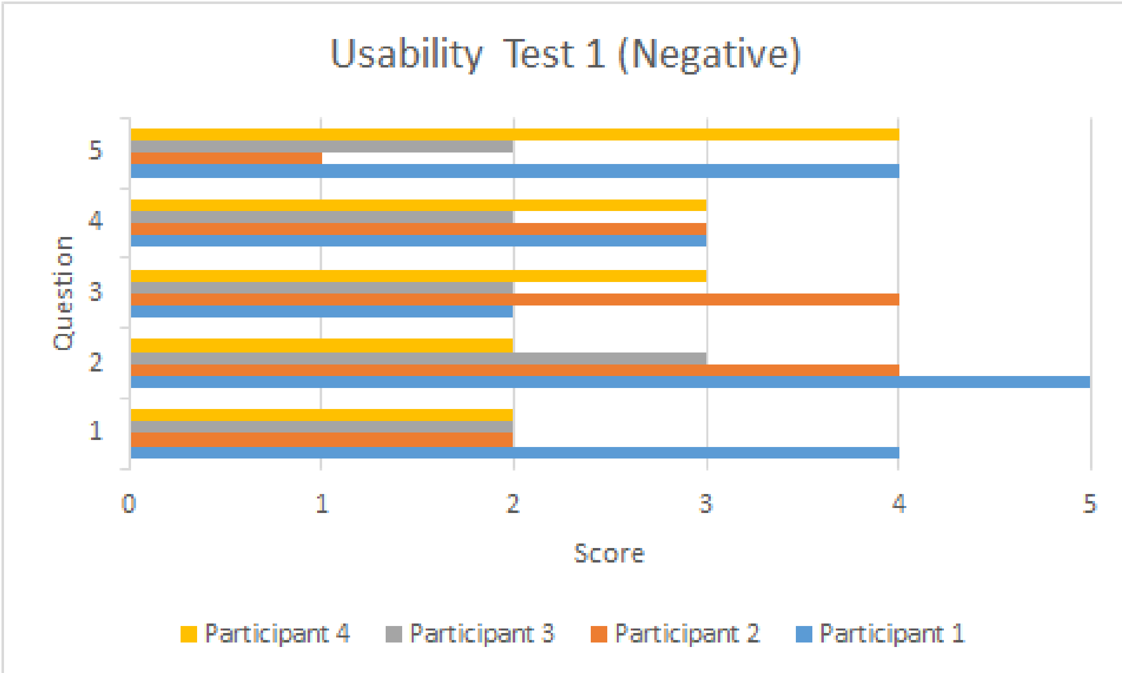
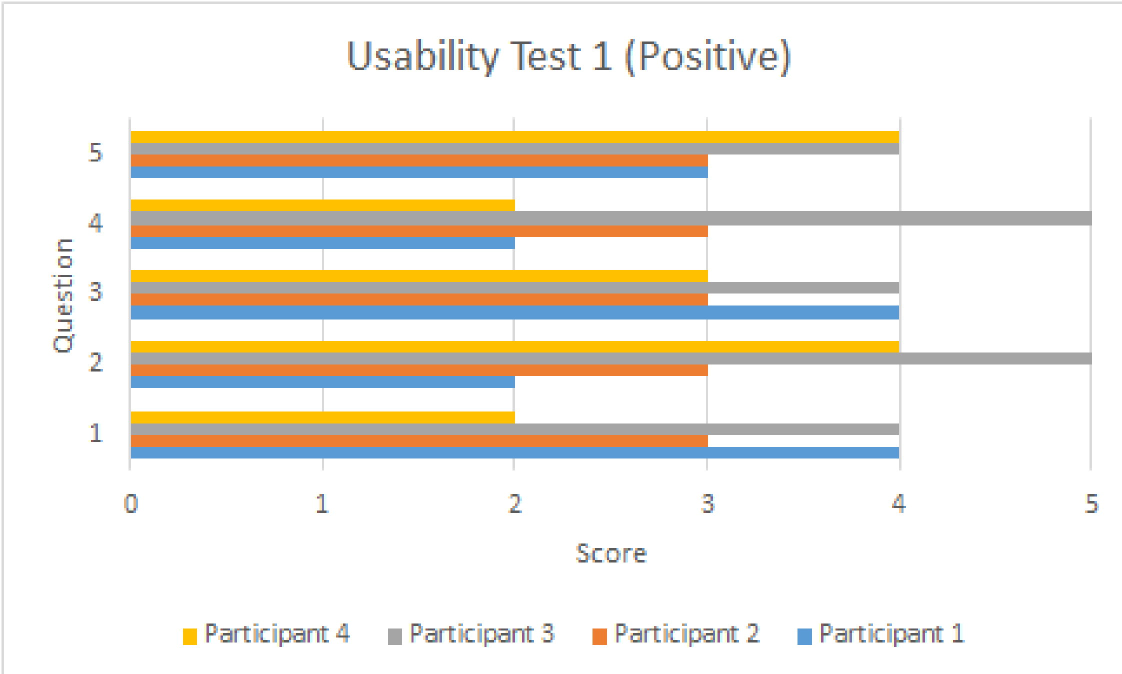


Figure 3: Graph of System Usability Survey positive prompts for first study



Plenty of written, verbal, and other unstructured feedback was also obtained that proved useful for informing future design decisions or identifying and fixing bugs. Some issues people had were as follows:

Figure 4: Issues from first study

- Not easy to go back and change accidental input during setup
- Unclear how to remove unwanted alarms (instead of just turning them off)
- No way for users to change their sleep goal after setting it the first time
- Scrolling through months to pick birthdate is inefficient
- Alarm did not have sound or did not play the chosen sample sound
- There are no clear instructions
- When screen falls asleep, it starts at the beginning screen again
- Main page top left corner: left facing arrow not a "go back" button
- 'finish' button in setup step didn't always work, sometimes crashed app
- App crashed after adding a new alarm or modifying an alarm
- App file from website couldn't install on 'Sharp Aquos Crystal' phone
- App file from website couldn't install on 'HTC One M9' Android phone

The participants also felt the following features would be useful tools for managing sleep health:

Figure 5: Recommendations from first study

- A feature that tracks how deep sleep is, as well as total sleep time
- Options for naps, bimodal sleep patterns, etc. in sleep goals
- A feature to calculate sleep adjustments for travel plans across time zones to help avoid jet lag.

The instructions in the focus group script were found to be confusing in a couple places: the "sleep scale" it mentions is called the Epworth Sleep Scale in the app, and setting sleep goals happens as part of the setup, which should be made more clear. The app ran properly on Android Motorola (Moto X), Nexus 5 and Nexus 6. For the next focus group it would be useful to collect the type of Android phone people used, to track any additional bugs and compatibility issues.

In our second usability test two focus groups were organized containing 3 and 2 WPI students respectively, and one student who participated individually for a total of 6 students. One of the participants was a graduate student at WPI, and the rest were undergraduates. The group consisted of 4 women and 2 men, one Computer Science major, two Industrial Engineers, two in Information Technology, and one in Management and Information Systems. As with the previous study, the groups did not all focus on one flaw or feature, and asked for help when they ran into an issue or were confused. The same usability metrics were recorded as in the first study, but notes were taken on the types of phones to help with tracking bugs in case they were encountered. There was an HTC_One_Max running Android 5.0.2 and a Samsung Galaxy S6, which ran without bugs or errors when installing. Also used were a Samsung Galaxy S5, a Samsung Galaxy S3, and a Oneplus One Android phone, which each had a little trouble from the same error interfacing with the Google Play store.

The expectation measure from Albert and Dixon [1] was recorded the same way as in the first usability test. The results were as follows:

Table 3: Expected vs. actual difficulty levels in second study

Difficulty	user 1	user 2	user 3	user 4	user 5	user 6
Expected	1.5	4	7.5	3	5	1
Actual	5.5	5	8.5	3	7	2

As in Study 1, it was again recorded in a task success matrix, though the task of setting a sleep goal was merged with the initial setup in the app for this round of testing. For more details see the changes made to the focus group script in appendix B. The results are compiled in table 4:

Table 4: Success rates of tasks in second study

Subject	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
1	1/2	1	1	1	1	1	1
2	1/2	1	1	1	1	1	1
3	1/2	1	1	1	1	1	1
4	1/2	1	1	1	1	1	1/2
5	1	1	1	1	1	1	1/2
6	1	1	1	1	1	1	1

Where the tasks in table 4 are drawn from the focus group script in appendix B. They are mostly the same, but with some clarifications and adjustments according to changes in implementation. The new tasks can be found in the following table:

Figure 6: Focus group 2 tasks

1. Get Android app from:
   https://play.google.com/apps/testing/sleephealthiqp.wpi.edu.sleephealth
2. Proceed through the initial setup procedure of the app, set sleep goal
3. Set a time for an alarm
4. Set music for an alarm
5. View the graph of sleep data
6. Take the Epworth Sleep Scale Survey
7. Opt out of notifications/alarms

The same System Usability Scale was used in the second study as the version used in the first study. The results of that are as follows.

Figure 7: Graph of System Usability Survey negative prompts for second study

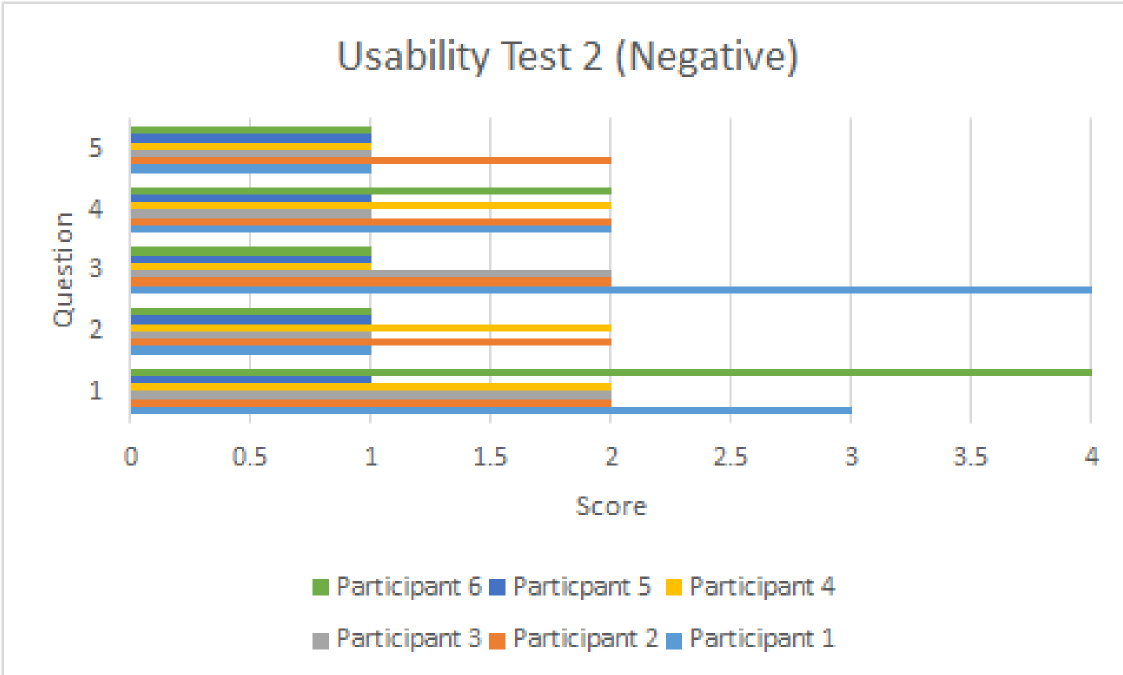
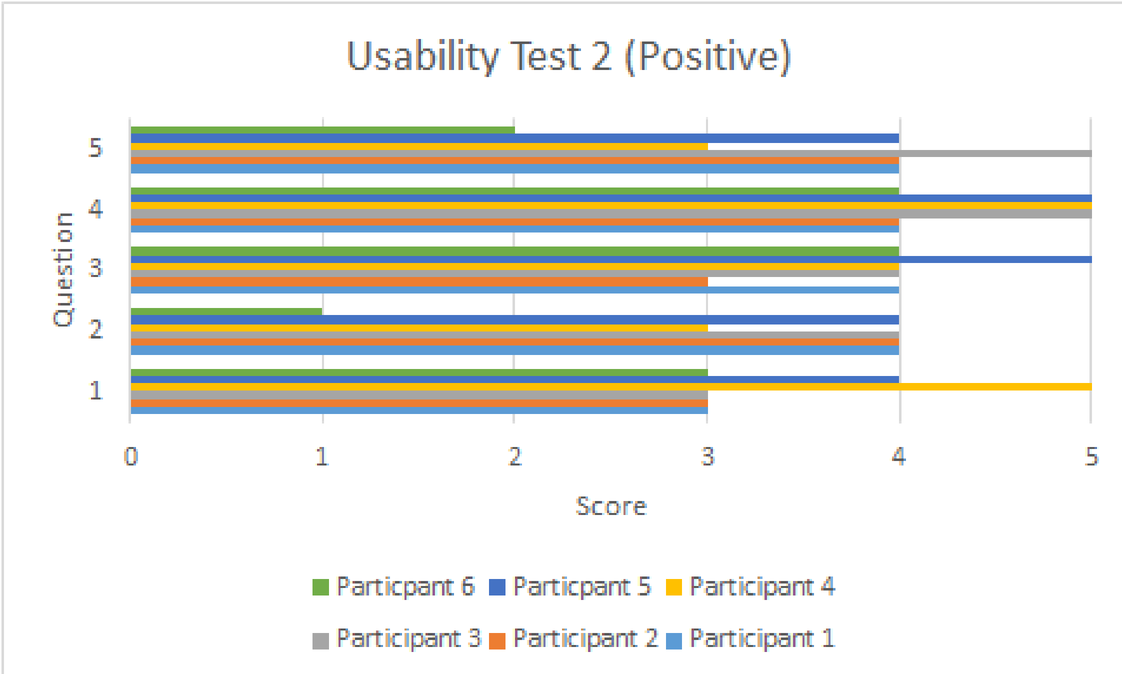


Figure 8: Graph of System Usability Survey positive prompts for second study



The second usability study also got useful verbal and unstructured feedback. The remaining issues encountered are as follows:

Figure 9: Issues from second study

- Setting the birth date by scrolling through the months is inefficient and mildly annoying
- The Button to set alarm sound is hidden at page bottom, same color as background
- Instructions are helpful, but the app could use more
- For some phones, Google Play says the app is not available at first: need to click again
- Still can't remove alarms, reminders, etc easily: gets cluttered eventually
- On Galaxy S5, certain buttons did not animate on button press: unclear whether they worked
- Bright blue text on a black background can be hard to read for some people
- Need to be able to change setup info if mistakes are made/sleep goals change

The participants also provided the following suggestions for improvements to the app:

Figure 10: Recommendations from second study

- calculate recommendations of when to sleep to adjust sleep cycle to goals with least lost sleep
- Add support for alternative sleep goals like bimodal sleep or naps
- Suggestions, info and advice on demand, as well as built into notifications
- The clock format for setting alarms is nice but could be done in less button presses
- keep sleep statistics like average sleep per week for the user to look at, along with graphs

7 Discussion

The participants of the first study were all confident Android phone users, but the participants in the second study were not. This made the results of the studies all the more encouraging, because despite the second group having somewhat less experience with Android phones, most of the quantitative metrics collected suggest that it was easier to use. The task completion percentage for Study 1 had an average of 68.75% with a standard deviation of 15.3%, and for Study 2 it is 92.86% with a standard deviation of 4.5%, following the removal of bugs and the improvement of the interface. The SUS scores demonstrated an encouraging improvement in usability as well. From a set of around 2000 SUS scores in 200 studies between 1996 (when the SUS score was developed [3]) and 2015, the average SUS score was 70.91 out of a maximum of 100 [2]. The set of SUS scores for the first usability test had a problematic average of 56.25, but with all the feedback from the first study, the score was able to be improved to a very acceptable 76.66. Further improvements can be made, as noted in the results of the second survey, but as it stands this app should provide a stable foundation for the exploration and testing of behavior change support systems. It is worth noting that objective usability at 92.86% was much higher than the SUS score of 76.66 (not a percentage, but the SUS is out of 100 anyway), and yet still 76.66 is a perfectly acceptable average SUS score. Clearly, usability is more utilitarian than ease of use or ergonomics. The average experienced difficulty minus expected difficulty was 1.5 for Study 2, but was only 1.0 for Study 1. There was however a lot of variability with this metric. The standard deviation of experienced minus expected difficulty in Study 1 was 2.345 and for Study 2 it was 1.95789. These are both at least 3 times larger than the difference between the results of the two studies, so from this metric it cannot be said with confidence which version was perceived as harder to use from this metric. The other quantitative metrics worked better at the scale of this small survey. If the trend continues to larger data sets though, it might suggest that the app is easier to learn for experienced Android phone users than for beginners, which would make some sense because the app drew some inspiration from other Android apps. During the second study it was clear enough that the app was in a usable state, because study participants did not even require help to install or use the app. Overall results were encouraging, and based on the data collected, it can be inferred that the Android app can be used successfully to test the robustness of BCSS and help WPI students improve their sleep health.

The app ran properly on Android Motorola (Moto X), Nexus 5, Nexus 6, HTC_ One_ Max running Android 5.0.2, and on Samsung Galaxy S6. Older phones like Samsung Galaxy S5, a Samsung Galaxy S3, and Oneplus One each had a little trouble from the same error interfacing with the Google Play store, but could at least still do it. Downloading an Android app off a website was much less viable than using Google play, because many Android phones (Like Sharp Aquos Crystal and HTC_ One_ M9) are not configured to recognize files as .apk when downloading them off websites, and therefore cannot install apps downloaded from the web. This is why so few people in the first focus group could download and install the app properly in the first place. In general the older or less standard phones were more likely to have compatibility issues, and moving forward to future projects it would be valuable to keep track of those issues. There were some nonspecific reports of buttons that did not animate when they were pressed and alarm sounds not playing. These were too intermittent to isolate when debugging, but they might show up again. All reported bugs in areas critical to the functionality of the interface were all fixed, though there may be a few minor issues to track down eventually.

The app usability is already reasonable, but significantly more feedback was obtained than could be incorporated into the app in the amount of time that was available. This leaves clear paths forward for future teams that want to further improve the app. Most important are the features and changes people kept asking for - even across studies and in separate groups. These goals should be a higher priority to implement, when the time comes to add additional features. Multiple study participants felt that scrolling back through the months to input their birth date in the app setup was inefficient and could be made more straightforward, and many participants wanted to modify or delete various data like inputs to the setup, alarms, reminders or accidental survey responses. Two study participants also wanted support for different types of sleep goals, like bimodal sleep cycles and naps. Some of the users also asked if the app could be used to mine its own data not just to provide suggestions, but to provide practical functionality to the user. For example, if the user has a flight to another country and wants to avoid jet lag, they could ask the app to construct recommendations based on their sleep patterns. This same algorithm could also provide the user with specific recommendations on when to go to bed each night to adjust to a new sleep goal. In another use case, the app could display the mean, median mode and range of the user's sleep data along with the graph of sleep data. This feature would enable self-tracking behavior for both visual people and analytical people.

Though not as universally in demand, individual participants also thought of some clever improvements to make. The button to set the alarm sound was also at the bottom of its page, and the same color as its background, so it was hard for one participant to find. Instructions and guidance built into the app proved very useful to study participants, but one participant asked if even more hints could be added. Another participant could not read the blue text well on the black background, and asked if they could change the app background color, but that feature was commented out in the code base at our advisor's request. One user wanted a feature that tracks how deep sleep is using the accelerometer, like the sleep clock app. It was originally thought constantly checking the accelerometer without skilled optimization would be a big drain on battery power, but if it only measures that information when the user is asleep and the phone is presumably charging, then it won't matter if it's a bit battery intensive. This was realized far too late through the IQP to try implementing that functionality, but it would make a great feature for future teams to add. One participant asked if the alarm and sleep goal setting method could be done in less button presses, though most people really liked that feature. Another participant asked if instead of notifications being a bothersome pop-up like thing on the Android phone, there could be suggestions, info and advice provided by the app on demand. This way if the user is curious or would like help or advice, they can get it themselves instead of having notifications bother them throughout the day. Curious users would likely be more receptive to advice. Making sure to pique the user's curiosity so that they will actually look at recommendations might be a challenge though.

8 Limitations and Future Research

The two studies used small, convenience samples of Android phone users who are students at Worcester Polytechnic Institute, and as such do not have enough data to perform sophisticated statistical analyses. However, these are formative studies to check the user experience, sanity-check our design decisions, and find bugs in the code to fix. Many of the early issues in development are clear enough to see without statistical analysis. As for examining the robustness of BCSS theory, getting the user interface and critical code structures to an easy-to-use state that won't crash comes first. In order to test overall efficacy, a longitudinal study over a long period of time is necessary, with larger study populations and more elaborate statistical techniques, so that task will likely go to a future IQP team.

In terms of future features to implement, many user recommendations to potentially implement were mentioned or analyzed in the discussion. There has also been much discussion amongst our own team about what ought to be implemented in the future. It would be nice to collect data without requiring the user's attention, like measuring acceleration over the course of sleep to gauge sleep quality, or measuring sound and light to evaluate whether the sleep environment is excessively noisy or bright. But such features may take up battery life and clutter the phone with background tasks, so to be properly implemented they should be intermittent, very efficient, or the app should remind the user to plug in their phone before going to bed. Also once the app collects its data, it doesn't really do much with it yet:

as of yet there is no algorithm which decides what advice to provide based on the collected data. It would be a good future experiment to try a number of different advice-choosing algorithms and see which are most effective. Data analysis could serve as a powerful support system for behavioral change, because it can allow the user to understand and monitor their sleep health in a way they might otherwise find too difficult or tedious, providing perspective they wouldn't normally consider.

Of particular interest is the development of community features: social support structures fit into BCSS theory and may be a good way to address the problem of poor sleep health in colleges, since students view their lack of sleep as a shared, social part of the college experience. Developing community features also allows us to more effectively incorporate the BCSS concepts of "rewards" and "similarity" into our user experience, because users could congratulate each other for developing good sleep related behaviors, compete for developing healthier behaviors in a manner similar to what Fitbit enables, or even just know that there are others out there confronting the same issues. In one future implementation, picture having users give advice about what solutions they find for fixing their sleep health, gaining a score based on how useful others find their solutions, effectively "gamifying" the search for practical sleep solutions. A filtered set of these recommendations could be distributed to people with similar sleep situations using the recommendations functionality of the app, helping to build interactivity and community aspects into a structure like a support group. Social features also allow the app to promote similarity as a persuasive tool, in a similar manner to the persuasive techniques tested in Lu's paper [10].

There is also no reason future versions of the Sleep Health app couldn't be modified to address other unhealthy or unproductive behaviors besides sleep health. Stripped of its sleep specific information the app could serve as a good foundation to build other BCSS-inspired programs.

9 Conclusion

Our contribution to practice is a stable, usable Android app aimed at improving the sleep health of students at WPI. We have enabled contribution to theory by constructing the app with Behavior Change Support Systems, in order to test their robustness when applied to real world problems. Initial development of the Sleep Health app using Behavior Change Support Systems and User Experience Design concepts went well, with all features of interest implemented. After usability testing with a group of 4 participants, the app was debugged and refactored to ensure that the final version of the application would be able to serve as a strong foundation for future testing and development of Behavior Change Support Systems. Subsequent usability testing with 6 students generated lots of valuable feedback for future refinements of the app, and quantified the improvement from the first round of app refinement, demonstrating good overall usability with above average System Usability Scale scores and a task completion rate of 92.86%. While the app developed during this IQP is fully functional, future research should focus on testing how effective the features are in promoting behavioral change. Implemented features include alarms, sleep tracking, and surveying to promote self-monitoring and reminders. A database of information is collected which can be analyzed to personalize the user experience in an automated fashion or provide the user with useful recommendations based on their data. These features can support behavioral change, and were designed with the intent of promoting healthy sleep behavior amongst WPI students. The code used for this project was documented heavily, and further information can be found in the technical document for the project, so future teams will be able to go in and change various aspects of the app structure in order to expand upon the existing project. As the initial programmers and researchers for this IQP, it will be exciting to see it expand into a more extensive study into how technology can be used to affect human lives for the better.

10 Bibliography

References

[1] William Albert and E Dixon. Is this what you expected? the use of expectation measures in usability testing. In *Proceedings of the Usability Professionals Association 2003 Conference, Scottsdale, AZ*, 2003.

[2] Aaron Bangor, Philip T Kortum, and James T Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6):574–594, 2008.

[3] John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[4] Mary. Carskadon, Thomas. Roth, Ruth. M Benca, Ronald. E Dahl, William C. Dement, Mark. Mahowald, Jodi A Mindel, Kyla. L Wahlstrom, Amy. R Wolfson, Susan. D Sagusti, Megan. Spokas, and Richard. Gelula. Adolescent sleep needs and patterns: Research report and resource guide. NIH funded research report, 2000. .

[5] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A Kientz. Sleeptight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 121–132. ACM, 2015.

[6] B. J. Fogg. *Persuasive Technology: Using Computers to Change What We Think and Do.* Morgan Kaufmann Publishers, 2003.

[7] Marc Hassenzahl. The thing and i: understanding the relationship between user and product. In *Funology*, pages 31–42. Springer, 2005.

[8] Sitwat. Langrial, Harri. Oinas-Kukkonen, Paivi. Lappalainen, and Raimo. Lappainen. Influence of persuasive reminders and virtual rehearsal on information systems for sleep deprivation. In *PACIS 2014 proceedings*, page Paper 228. Association for Information Systems Electronic Library.

[9] James R Lewis and Jeff Sauro. The factor structure of the system usability scale. In *Human Centered Design*, pages 94–103. Springer, 2009.

[10] Amy Shirong Lu. An experimental test of the persuasive effect of source similarity in narrative and nonnarrative health blogs. *Journal of medical Internet research*, 15(7), 2013.

[11] Jakob Nielsen. Usability inspection methods. In *Conference companion on Human factors in computing systems*, pages 413–414. ACM, 1994.

[12] Harri Oinas-Kukkonen. A foundation for the study of behavior change support systems. *Personal and ubiquitous computing*, 17(6):1223–1235, 2013.

[13] Harri Oinas-Kukkonen and Marja Harjumaa. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1):28, 2009.

[14] Kathryn M Orzech, David B Salafsky, and Lee Ann Hamilton. The state of sleep among college students at a large public university. *Journal of American College Health*, 59(7):612–619, 2011.

[15] Tom Tullis and Bill Albert. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. 2013.

[16] Vance Wilson and Soussan Djamasbi. Human-computer interaction in health and wellness: Research and publication opportunities. *AIS Transactions on Human-Computer Interaction*, 7(3):97–108, 2015.

A Focus Group 1: Questions/Script

The following is a script of the questions and tasks given during the first focus group:

Before the testing:

- How old are you?
- What is your major here at WPI?
- Are you experienced using Android phones?
- You will be working with an app intended to improve student sleep health by interacting with the user. How difficult would you expect such an app would be to use on a scale of 1 to 10 where 1 is very easy and 10 is very difficult?

Tasks to attempt:

- Setup:
  - Get the Android app off of arc30.wpi.edu/sleephealth
  - Proceed through the initial setup procedure of the app
- Alarms:
  - Set a time for an alarm
  - Set music for an alarm
- Feedback:
  - Set a target sleep goal
  - View the graph of sleep data
  - Take the sleep quality survey

- Settings:
  - Opt out of notifications/alarms

Post questions:

- On a scale of 1 to 10 where as before 1 is very easy and 10 is very difficult, how hard was it to use the app?
- What features would you like to see in a sleep health app?
- What parts, if any, could be done better/differently?

B Focus Group 2: Questions/Script

The following is a script of the questions and tasks given during the first focus group:

Before the testing:

- How old are you?
- What is your major here at WPI?
- Are you experienced using Android phones? What model of Android phone do you use?
- You will be working with an app intended to improve student sleep health by interacting with the user. How difficult would you expect such an app would be to use on a scale of 1 to 10 where 1 is very easy and 10 is very difficult?

Tasks to attempt: (Put a checkmark next to successfully completed tasks, a dash if you got it right with some help, and make note of significant difficulties/points of confusion)

- Setup:
  - Get Android app from: https://play.google.com/apps/testing/sleephealthiqp.wpi.edu.sleephealth
  - Proceed through the initial setup procedure of the app, set target sleep goal
- Alarms:
  - Set a time for an alarm
  - Set music for an alarm

- Feedback:
  - View the graph of sleep data
  - Take the Epworth Sleep Scale survey
- Settings:
  - Opt out of notifications/alarms

Post questions:

- On a scale of 1 to 10 where as before 1 is very easy and 10 is very difficult, how hard was it to use the app?
- What features would you like to see in a sleep health app?
- What parts, if any, could be done better/differently?

C System Usability Scale

System Usability Scale

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

D Epworth Sleep Scale

The Epworth Sleepiness Scale

The Epworth Sleepiness Scale is widely used in the field of sleep medicine as a subjective measure of a patient's sleepiness. The test is a list of eight situations in which you rate your tendency to become sleepy on a scale of 0, no chance of dozing, to 3, high chance of dozing. When you finish the test, add up the values of your responses. Your total score is based on a scale of 0 to 24. The scale estimates whether you are experiencing excessive sleepiness that possibly requires medical attention.

How Sleepy Are You?

How likely are you to doze off or fall asleep in the following situations? You should rate your chances of dozing off, not just feeling tired. Even if you have not done some of these things recently try to determine how they would have affected you. For each situation, decide whether or not you would have:

- No chance of dozing =0
- Slight chance of dozing =1
- Moderate chance of dozing =2
- High chance of dozing =3

Write down the number corresponding to your choice in the right hand column. Total your score below.

Situation	Chance of Dozing
Sitting and reading	•
Watching TV	•
Sitting inactive in a public place (e.g., a theater or a meeting)	•
As a passenger in a car for an hour without a break	•
Lying down to rest in the afternoon when circumstances permit	•
Sitting and talking to someone	•
Sitting quietly after a lunch without alcohol	•
In a car, while stopped for a few minutes in traffic	•

Total Score = ________________________

Analyze Your Score

Interpretation:

0-7: It is unlikely that you are abnormally sleepy.
8-9: You have an average amount of daytime sleepiness.
10-15: You may be excessively sleepy depending on the situation. You may want to consider seeking medical attention.
16-24: You are excessively sleepy and should consider seeking medical attention.

Reference: Johns MW. A new method for measuring daytime sleepiness: The Epworth Sleepiness Scale. *Sleep* 1991; 14(6):540-5.