UNIVERSITY OF TWENTE.

Faculty of Biomedical Engineering,
Biomedical Systems and Signals

Investigating technical aspects, usability aspects, and participatory willingness for an effective mobile data sensing study

Monique te Rietstap
MSc. Thesis
January 2018

Graduation committee:
Dr. Ir. B.J.F van Beijnum
Dr. Ir. O. Banos
Prof. Dr. Ir. H.J. Hermens
Prof. Dr. H.K.J. Heylen

Telemedicine Cluster of the Biomedical Signal and Systems Group
Faculty of Science and Technology (TNW)

INVESTIGATING TECHNICAL ASPECTS, USABILITY ASPECTS, AND PARTICIPATORY WILLINGNESS FOR AN EFFECTIVE MOBILE DATA SENSING STUDY

BY

MONIQUE TE RIETSTAP
S1716891

THESIS

Submitted in partial fulfillment of requirements for
the degree of Master of Science in Biomedical Engineering
University of Twente, 2018

Enschede, the Netherlands

Adviser:

Dr. Ir. Oresti Banos

Dr. Ir. Bert-Jan van Beijnum

Abstract

It is impossible to imagine a world without smartphones in this day and age. Moreover, the development of new technologies involving smartphones is occurring at a fast pace. It is common knowledge, that most mobile devices are equipped with sensors that collect data such as body motion (via accelerometer) and whereabouts (via GPS or Wi-Fi). It is possible to use these sensors to continuously monitor daily behavior, which is a technique referred to as mobile data sensing. One such exciting research technique is the AWARE framework. This framework can collect implicit and explicit user-mobile interactions from smartphones anonymously. Despite the promising outlook, little research has been done on the willingness of subjects to participate in mobile data sensing studies.

Stress is a phenomenon experienced by most students. When exposed to stress for long periods of time it can be a serious strain on both physical and mental health. A solution for this could be to continuously monitor stress through mobile data sensing. In this thesis, a larger project is kept in mind in order to conduct the experiments. This is to study stress and explore the possibility to derive an empirical model for stress using data collection from smartphone usage data. Before it is possible to conduct such a large-scale experiment it is necessary to evaluate and understand the following objectives. To investigate the usability and technical aspects of a system such as AWARE. In addition, to investigate the willingness of students to participate in mobile data sensing studies. Three experiments were conducted to investigate the two objectives, two to fulfill the first objective and one to reach the second objective.

The first experiment consists of an observational study in which a mobile data sensing study was introduced, to the reader, using sensors that were relevant to stress. Students joined this study by installing the AWARE (mobile data sensing) and Paco (survey app) to ask questions throughout the day. The obtained data was used to investigate the usability of the experiment, the response of participants to the survey app that asked questions throughout the day, and jitter in the sensor data. The second experiment was conducted in a controlled environment in which students installed the AWARE app and carried the smartphone containing the app around for 48 hours. The collected data was used to investigate battery drainage caused by AWARE, and possible jitter in the location tracking sensor data. In addition, the installation process was discussed with the participants. In the third experiment, interviews and an online survey were conducted to investigate the willingness of subjects to participate in mobile data sensing studies. The topics used in the interviews and online survey included: privacy, security, participation motivation, perception towards the different mobile sensor data, length of experiment, reward, being able to see your own data, and possible suggestions of participants.

The results show that minor jitter was found in the sensor data which suggests that the sensor data collection is consistent. AWARE drains the battery approximately 40 to 60% over a 24-hour period, indicating that smartphones need to be charged regularly to ensure consistent data collection. This entails charging the smartphone at least once a day. Data collection in terms of calls, texts, notifications, and device use have been collected without any data gaps (i.e. no missing data) to the extent of our knowledge. In addition, the students who participated thought the experiment was easy in use. On average the response time to the questions was 5.30 minutes and 74% of all questions were answered. Some suggestions were given to improve a future installation process of AWARE. Lastly, the interviews and surveys resulted in some factors that determine the interest users have in participating in a mobile data sensing study. These factors include an experiment duration of three to five weeks, ESM with a frequency of two to three times a day, a reward for finishing the experiment, let participants choose data to be measured from a list, and recruit prospective participants with a motivating pitch.

Acknowledgements

I would first like to express my gratitude to my daily supervisor Dr. Banos of the BSS faculty at the University of Twente. His door was always open whenever I needed advice and also provided me with beneficial remarks and useful insights throughout this incredible learning process. Furthermore, I would like to thank to thank Dr. Ir. van Beijnum for introducing me to this new topic and for the continuing support and advise along the way. In addition, I would like to thank Dr. Noordzij for providing me with great advice proceeding with my last experiment. I would like to acknowledge Dr. Heylen and Dr. Ir. Hermens for their participation in my defense committee. I also like to acknowledge the subjects who participated in my experiments for their valuable insights and precious time.

Finally, I'd like to express my profound gratitude to my loved ones for their unconditional love and support throughout this entire process. Thank you.

Table of content

List of Figures

List of Tables

Introduction

This chapter provides the reader with an introduction to this area of research. It presents the motivation, main objectives, and sub-objectives. In addition, the approach employed in this thesis and outline is presented to guide the reader through the thesis.

1.1 Background and motivation

Nowadays, mobile devices are ubiquitous and are equipped with various sensors. Often smartphones are able to sense body motion (via accelerometer), noise (via microphone), whereabouts (via GPS or Wi-Fi), and a lot more. These sensors have the ability to collect large quantities of data which that can be used for various purposes, such as physical or mental health-related research. This technique is referred to as mobile data sensing [60]. Collecting data continuously through mobile data sensing can provide detailed insides into the lives of people. This makes for a very exciting research technique, one example of this is called the AWARE framework.

The AWARE framework is an Android-based framework committed to log, share, and collect data from smartphone usage. [35]. The AWARE framework is an application that uses the built-in sensors of a mobile device and collects data from the smartphone use of people. This app is used as a research technique to analyze data by visualization or data minding. This app on a smartphone is able to measure and collect various types of data using sensors (readily available in an Android phone) and plugins (e.g. more advanced sensors), such as hardware (i.e. accelerometer) and software sensors (i.e. logging of calls and messages). The sensors and plugins serve the purpose of observing and monitoring collected data, the application makes it possible to find relations between different sensors of context information, such as social and individual behavior. Additionally, AWARE is a very special application as it ensures complete anonymity of the end-users, by encrypting every information transfer route and giving IDs to the mobile device and specially encrypted IDs to source information of calls and messages [32].

Stress is present in the daily lives of students. Stress is experienced due to homework, projects, exams, social obligations, and more. A bit of stress is not bad. However, when exposed to stress for longer periods of time, it can have serious consequences for both mental and physical health [34]. Assessing stress in students is therefore very important to keep them as healthy, which is both important for students and university. Current measurements of stress are conducted sporadically in lab settings and with self-report. A consequence of sporadic measurements is that important information can be missed, meaning that little known about stress throughout daily life [31]. Fortunately, a solution to stress monitoring can be mobile data sensing as it can measure mobile user data continuously through mobile data sensing. There is a large goal kept in mind to conduct the experiments in this thesis. This is to study stress and explore the possibility to derive an empirical model for stress using data collection from smartphone usage data. Before it is possible to conduct a large-scale experiment, it is necessary to evaluate and understand the technical aspects (and usability) of a system such as AWARE. Investigating these two aspects will make it possible to conduct future mobile data sensing studies properly and without complications. In addition, the willingness of students to participate in such an experiment needs to be assessed. The reasons for this is that little research has been conducted on the willingness of people to participate in mobile data sensing studies.

1.2 Research objectives

In this section, the two main objectives will be described and explained. The two main objectives are as follows.

- To investigate the technical and usability aspects relevant for an effective mobile data sensing study
- To investigate the willingness of subjects to participate in mobile data sensing campaigns.

First, the technical aspects that will be investigated are for example the of types of sensors that are relevant to stress, to analyze the collected data from the mobile data sensing study, to analyze the design of the mobile data sensing study, and to improve the design of the experiment where necessary. Secondly, usability aspects are examined to analyze the design of the mobile data sensing study with respect to the experience of the participants. This entails, how have participants experienced the mobile data sensing, how can the experience be improved, and what aspects of the mobile data sensing study are experienced as pleasant. Lastly, the users' willingness to participate in a mobile data sensing study will be investigated. Aspects investigated are for example motivation to participate, the users view on different sensor types, and the participants view on being monitored via the users' smartphone continuously throughout the day.

1.3 Research approach

This research is divided into three experiments. Figure 1 shows the research model of this thesis. It shows how the theoretical foundation is connected to the experiments performed and how the experiments are linked to the objectives of this master thesis. In what follows the approach employed to obtain the research objectives will be discussed.

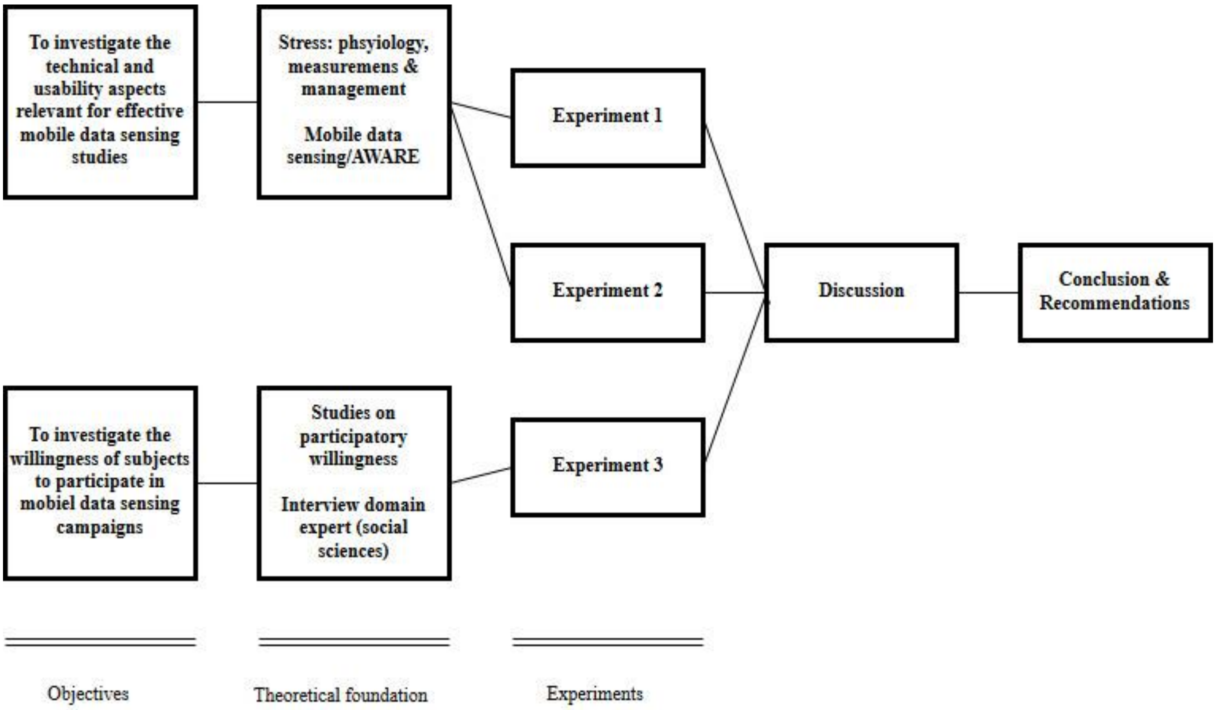


Figure 1: Research model

A literature study is performed to describe the theoretical body of this research. The topics with respect to mobile data sensing, physiology of stress, and state-of-the-art of stress management and measurement. In addition, the literature will be studied to investigate the willingness of students to participate in mobile data sensing studies.

Furthermore, an interview with a domain expert (social behavior researcher) will be conducted about the participatory willingness in mobile data sensing studies. It will be discussed what the best method of approach will be to assess the participatory willingness, and aspects will be discussed such as recruitment, survey questions, and experiment length.

Experiment 1: The first experiment is an observational and anonymous study in which a mobile data sensing study will be introduced, which uses sensor parameters that are relevant for monitoring stress. In this thesis, this study design will be referred to as ***mobile stress study***. This study uses the AWARE app for mobile data sensing and the Paco app (i.e. a survey app) to ask questions during the day. The apps run parallel throughout the length of the study. When the study is finished the subjects are sent a survey to investigate the how the study has been experienced by the participants. The data collected with AWARE and Paco will be analyzed using quantitative analysis in MatLab.

Experiment 2: The same mobile data sensing experiment is used as in experiment one, with the addition of a Wi-Fi sensor (to track location) and a battery sensor (to track battery levels). In this controlled study, only AWARE is used to assess the usability and experience of the installation process. Students were asked to install the AWARE application on smartphones and to participate in the mobile data sensing experiment for 48 hours. After the installation process, a face-to-face evaluation of this process was discussed. The collected data after 48 hours was used to analyses. This quantitative analysis is performed in MatLab. This is conducted in a second experiment as these aspects were considered after the first experiment was performed.

Experiment 3: Combining the literature and the interview with a domain expert results in the design of experiment 3. In this experiment, an interview and an online survey are designed to investigate the willingness of students to participate in mobile data sensing studies. For this experiment, a modified version of the mobile data sensing study (***mobile stress study***) is introduced to the subjects of experiment 3. This modified version is the one that will be used to reach the larger project (mentioned in 1.1). This is to derive an empirical model for stress using data collection via mobile data sensing. In the interviews and surveys, it will be investigated which factors contribute to the willingness of students to participate in mobile data sensing studies. The topics that were considered of importance to investigate for the willingness included:

- Perception towards the different mobile sensor data;
- The length of the experiment;
- The design of the study;
- The perception towards question frequency and amount of questions at a time;
- Motivation to participate;
- Ways of recruitment;
- Understanding of consequences when participating in mobile data sensing experiments;
- Privacy, and security.

In addition, after the experiments have been conducted the results will be addressed and discussed in the discussion chapter. After which, the research objectives will be addressed in the conclusions and recommendations will be given to help aid in future mobile data sensing research.

1.4 Report organization

The report is organized into six chapters:

Chapter 2: Theoretical foundation: This chapter focusses on introducing the theoretical foundations for this master thesis, topics include: stress, state-of-the-art, mobile data sensing, and participation willingness of mobile data sensing studies.

Chapter 3: Experiment 1: In this chapter investigates if there is jitter in the sensor data, to investigate the response of participants with respect to the questions sent via the Paco application, and to investigate the usability of the AWARE app, the Paco app, and how the experiment was experienced by the subjects.

Chapter 4: Experiment 2: This chapter assesses the usability of the client, investigates the battery drainage caused by AWARE, and investigates if there is jitter in the location sensor data.

Chapter 5: Experiment 3: Investigates the willingness of subjects to participate in mobile data sensing campaigns.

Chapter 6: Discussion: Evaluation of results and comparing findings to literature.

Chapter 7: Conclusions and recommendations: Addressing the research objectives, sub-objectives, and making recommendations and guidelines for future research and for the larger goal mentioned in the introduction.

Theoretical foundation

This chapter presents the theoretical background of this research thesis. The topics include physiology of stress, current stress measurement and management methods, mobile data sensing, and the willingness to participate in mobile data sensing studies. The topics will be discussed and the state of the art is given. Moreover, an interview was conducted with a domain expert to get information on the willingness of students to participate in mobile data sensing experiment from a social science perspective.

2.1 Physiology of stress

The Paleolithic man used stress as a survival mechanism. The ability to feel stress made the cavemen attentive to a potential hazard. The primary response to stress is physical, the body goes into a state of 'fight-or-flight'. This response leads to release of 14biochemical and hormones (i.e. cortisol, and adrenaline) into the bloodstream, causing physical reactions such as increased blood circulation to provide muscles with more oxygen, increased heart rate, and decreased body functions that are futile at that particular time. These responses provide the man with an energy charge, it supplied the cavemen to either fight or flees during a dangerous situation.

Nowadays, the same 'fight-or-flight' response is still being used in, for example, dangerous situations or situations of anxiety (e.g. public speaking). It can provide the body with a:

- Flight response, this is a natural reaction to some situations in which a person could become aggressive or agile, a helpful response in praetorian situations. However, it could also have a negative effect on (social) affairs.
- Flee response, this is a bodily response by removing oneself from an alarming situation. It could be helpful to respond in this manner, however, if the stressor does not go away it will keep increasing the stress level.
- Freeze response, this is when the energy released during a stressful situation gets trapped in the nervous system. The body responds by freezing.

These three stress responses minimise the brain function as an increasing amount of blood is pumped through the muscles, which can impede our capacities in everyday situations. If the body stays in a state of stress for long periods of time, it can influence physical and mental health [1].

2.1.1 Chronic stress

As stated above, experiencing stress for periods of time can have devastating effects on both mental and physical health. This is referred to as chronic stress. In this state, an overexposure to adrenaline and other stress hormones is occurring. It could augment several health risks such as anxiety, depression, headaches, weight gain, and insomnia. These health risks can be aggravated when people try to deal with stress by smoking, overeating or other bad habits [2, 3].

There are a lot of stressors that could cause chronic stress, such as high workload, a high demanding job or challenging relationships. Everyone experiences these stressors differently and in some people, this leads to more or less stress than in others. This is said to be a result of genetics and life experience [4].

2.1.2 Biology: hypothalamic-pituitary-adrenal axis

A stressor is anything that throws the body of balance and instigates a stress response, examples of such stressors are: fear, infection or injury. The mechanism behind the stress response is called the hypothalamic-pituitary-adrenal axis or HPA axis for short. When a stressor is detected by the hypothalamus, the sympathetic nervous system causes the hypothalamus to release the hormone corticotrophin-releasing factor (CRF). This hormone activates the pituitary gland, which stimulates the production of adrenocorticotrophic hormone (ACTH), see Figure 2 [5], for the hypothalamus and pituitary gland. This hormone activates the adrenal glands (on top of both kidneys) to stimulate the adrenal cortex for the secretion of cortisol [6].

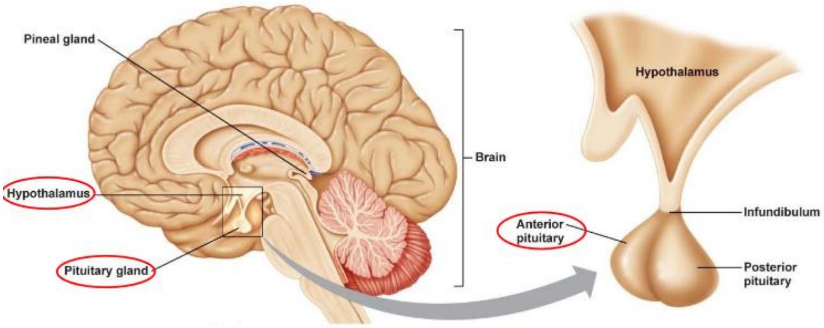


Figure 2: Hypothalamus and pituitary gland

Simultaneously, the neurons in the hypothalamus signal the medulla (inside the adrenal gland) to release epinephrine (i.e. adrenaline) and norepinephrine (i.e. nor-adrenaline). This brings the body into a state of hyper-alertness (fight-or-flight) [6].
When epinephrine and norepinephrine are released into the bloodstream, the blood vessels leading to the muscles, heart, and lungs will dilate. Subsequently, they constrict the smaller blood vessels as this allows for all the blood and oxygen to go to the major muscles. Another effect of (nor-) adrenaline is the reduction of pain, resulting in increased strength, performance and increased awareness.
Sometimes adrenaline is released when no real danger is present, which could result in for example dizziness. Additionally, glucose release takes place to increase the energy during hyper-alertness, however, this presents itself as being very dangerous when the glucose cannot be burned. It could lead to insomnia and heart damage.
If the stress continues the hypothalamus activates the release of glucocorticoid hormone cortisol, as the stored glucose has been used for immediate energy release. This hormone and a growth hormone secreted by the hypothalamus demand the body to release stored compounds from adipose tissue and muscle tissue to increase energy levels. In case of injury or infection, it provides nutrients to restore cells.
Cortisol and adrenaline will still be present in the bloodstream when suffering from chronic stress, which will lead to high blood volume, high blood pressure and a high glucose level in the bloodstream. Next, to that, cortisol also suppresses inflammatory and immune responses which can be very dangerous [7].

2.1.3 Stress and student life

Stress is a struggle of everyday life, this also includes students. Students must deal with new found freedom and responsibilities. Even students that have been at the university for a longer period of time have to deal with stress factors. Balancing these stress factors with school (i.e. high workload, projects, exams, etc.), holding down a job to pay for all the expenses, and social obligations (i.e. parties, friends, student associations) can be very difficult.
It could result in a lot of stress due to anxiety, pressure, and sleep deprivation. It can result in a decrease in both physical and mental health, causing some serious issues. The general health of students is very important for the university and therefore it is necessary to keep up their health. A way of doing this is by monitoring the stress of students as can give a proper indication of the mental and physical status of students. By observing stress in students, the university could, for

example, be advised on special schedules or training specially designed for specific students to keep their health optimal.

2.2 Current stress management and measurement methods

2.2.1 Stress management

Stress is part of modern life. The body cannot exclude the stress response. As it cannot be removed it is important to manage stress to relieve the body from a state of high alertness. If stress is not managed, over time it could result in serious health problems.

First, it is important to find what triggers the stress, this makes it easier to deal with the stress rather than performing random stress relieving techniques. According to an online poll by the ADAA (Anxiety and Depression Association of America), people deal with stress in descending order: social support from family and friends (18%), sleep (17%), exercise (14%), watching television (14%), eating (14%) and listening to music (13 %) [21]. A few of these relieving techniques will be elaborated on below.

Physical activity

In 2007 a study by Hansmann et al argued that performing different sports in the outdoors significantly reduces stress. Especially high-intensity sports (e.g. jogging and biking) helps more intensively with stress relieve as low impact exercise (e.g. walking or relaxing) [17]. Another study by Hamer et al has found preliminary evidence that suggests HPA-axis, automatic nervous system and other processes respond to exercise. This provides a link between mental well-being and exercise [18]. Moreover, quite a few studies have investigated the link between stress relieve and physical exercise [19, 20]. The studies presented here also associate a positive mood with increased physical activity.

The World Health Organisation (WHO), recommends 150 minutes of moderate exercise a week for adults between the ages of 18 and 64. This should be divided into intervals of 30 minutes 5 times a week. It only counts as a physical activity if the duration of the performed exercise has a minimum duration of ten minutes [22]. Walking, running and yoga are the preferred physical activities when relieving stress, according to an online poll of the ADAA [21].

It is said that exercise helps you to relax, it removes the tension from the body as the excess energy trapped in the body is burned off.

Social support

Social support has been described by the National Cancer Institute's Dictionary of Cancer Terms as: "A network of family, friends, neighbors, and community members that is available in times of need to give psychological, physical, and financial help" [24]. Theoretical models have been made of social support that is divided into:

- A category with a structural dimension, including frequency of social interaction and the size of someone's network
- A category with a functional dimension that includes instrumental components (e.g. gifts, financial help, and assistance) and emotional components (e.g. empathy and love).

According to research, functional dimensions are more important to health-related factors than structural dimensions [23]. Social support is very important for mental and physical health, it is argued that positive social support damps the activity of the HPA-axis, which results in a resilience to stress [23]. Social support, especially social interaction can be used as a mechanism to cope with stress and relieve it, as it reduces the levels of cortisol in the body that are released by the

HPA-axis [25]. Social support takes place between human beings, either face-to-face, over the phone or through social media.

Sleep

Enough sleep is one of the most important factors to function properly. The recommended amount of sleep by the National Sleep Foundation (NSF) is seven to nine hours for younger adults between the ages of 18 and 25 (e.g. students) [26]. When these recommendations are not met, the body does not experience all the benefits of sleep (e.g. revitalize the brain and body) [27, 28]. Even insignificant sleep deprivation could have health-related consequences such as bad mood and disinterest. Eventually, it could result in serious conditions such as high blood pressure [29, 30].
In 2013, the Stress in America™ survey suggested that stress might be impeding sleep. It states that the average adult in America receives about 6.7 hours of sleep a night and 43%redirects the deprivation of sleep to stress. Moreover, when the sleep quality decreases the number of stress increases.
When little sleep is received, the feeling of fatigue and exhaust increases. This makes it more difficult to focus on work, which can enhance the feeling of stress. Therefore, it is very important to get enough sleep, both for mental and physical health.
The methods of stress management mentioned above are the major methods for stress relief. However, there are also other methods to manage stress that is less scientific such as music and television.

Mood, happiness, and nutrition

Some studies have shown that stress is related to mood and happiness, this might be familiar to a lot of people get unhappy or grumpy when they have a lot on their plate or are very stressed. A study by DeLongis et al performed assessed 75 married couples for six months, they concluded that increased daily stress levels are linked to a disturbance in mood [44]. Another study suggests that an increase in perceived stress might be linked to a direct decrease in happiness. They suggest that intervention to reduce stress might be helpful to increase happiness [45].
It is well-known that proper nutrition is very important to be healthy and happy. In addition, research has shown that healthy nutrition has a positive effect on stress levels. Stress is thought of as an influencer to eating behaviors in humans. Overeating or undereating is associated with stress [46]. Healthy nutrition is important, especially breakfast. Breakfast is said to reduce the level of cortisol in the blood [47, 48].

2.2.2 Stress-measurement

Stress measurements can be performed in various ways such as psychological measurements and physiological measurements.

Psychological measurement

Psychology measures by focussing on questions. Since, the late fifties and early sixties, scientists have been developing self-reports. Psychologists have developed multiple surveys and questionnaires to detect various psychological symptoms that are triggered by chronic stress. One of the first and most well-known questionnaires on stress self-report is called the Social Readjustment Rating Scale (SSRS), which was developed in 1967 by Holmes and Rahe. The SSRS checks the stress ranking factors over the past six to 24 months and uses the results to see if there are correlations with future events such as disease. The relationship between this stress measure and such occurrence of illness or other factors can be determined [9].
Another stress test used often is developed by Cohen, et al. in 1983, and is called the Perceived Stress Scale (PSS). It is a scale with 14 items that pursuits the analysis of daily events that could

be perceived as “unpredictable, uncontrollable, and overloading” [8]. It evaluates a few aspects of stress:

- Important life events;
- Daily commotions;
- Alterations in coping abilities.

This scale has been tested during various experiments such as stress measurement with undergraduates during exam period, and the prediction of symptoms like fatigue and headaches in graduate students [10, 11]. The downside to these self-reports is the length of the surveys, often the questionnaire exceeds 30 questions.
For this project, it is necessary to use a shorter questionnaire that rates short-term perceived stress. An example for this is the Global Assessment of Recent Stress Scale or GARS scale for short. It has been developed to evaluate contemporary stress, additionally, it has been developed to evaluate immediate stress over a short-term (e.g. when investigating perceived stress and physiological tests simultaneously). This is different from other stress surveys, as most of these surveys capture the perceived stress by evaluating occurrences over the past 6 to 24 months.
Tests have been performed with respect to its factor structure, validity, and reliability. From the results, it was suggested that the GARS scale might be a useful tool to assess currently perceived stress by individuals [37].

Physiological stress measurement

When the HPA axis is activated various 18 biochemical are released into the bloodstream. As mentioned above, cortisol and (nor-) adrenaline are released by the hypothalamus. These substances can easily be measured by urine, blood, and saliva. Measuring cortisol levels in saliva has been a favorable measurement as it’s a non-invasive method and it does not require educated personnel to handle procedures. (Nor-) adrenaline, however, cannot be measured through saliva. It is measured by proxy measures, which is an indirect measurement of the parameter that is analyzed [15]. A few proxy measures are:

- Blood pressure: (nor) epinephrine increases blood pressure. It is a sympathetic parameter that measures systolic pressure and diastolic pressure. Systolic pressure is when the left ventricle contracts and exerts maximum pressure on the arterial walls. Diastolic pressure is the pressure when the heart is at rest. It can be measured with a blood pressure cuff.
- Alpha-amylase (Aas) is a substance that can be collected through saliva. Research by Li et al (2004) and Walsh et al (1999), suggests that increased Aas levels are related to increased stress [12, 13]. Subsequently, research by Chatterton et al has shown that Aas levels are correlated with levels of (nor) epinephrine [14].

In addition, markers such as ACTH and CRF levels can also not be measured through saliva but through blood and cerebrospinal fluid, respectively. During acute stress, vasoconstriction of the blood vessel just underneath the skin surface occurs. This leads to a drop-in skin temperature, which can be measured using infrared thermography. This analysis measures the change in skin temperature [16]. An overview of physiological measures of stress will be provided below with the accompanying measuring tests/ techniques [43]:

- Body gesture, using automated gesture analysis (leveraging AFEA);
- Facial expressions, using automatic facial expression analysis;
- Eye activity, using infrared to track the eyes;
- Blood activity, using PPG (photoplethysmography);
- Respiratory response, this uses electromagnetic generation;
- Muscle activity, using EMG;
- Brain activity, using EEG;
- Heart activity, using ECG,

- Also, skin response (i.e. skin conductance), using GSR and EDA.

Although collecting stress related data through the methods mentioned above is very scientific and accurate, the measurement of stress can only be collected in laboratory settings, surveys or self-reports. This results in several limitations such as the sporadic recordings of information that will not include specific details and can cause too much effort on the end-user's side for long periods of monitoring [31]. The lack of continuous measurement does not provide proper insight into the stress experienced by students throughout a day. To accumulate continuous data from a student, it is possible to collect data through mobile data sensing. The section that follows describes the monitoring of stress with mobile devices.

2.3 Stress and mobile data sensing

Nowadays, mobile devices are ubiquitous and equipped with various sensors (e.g. camera, accelerometer, and microphone). Smartphones make it possible to collect the data of mobile phone users, such as data from previously mentioned sensors, but also data from applications obtained from an app store. Moreover, continuous data sensing is possible, allowing more detailed insight into the lives of students. This makes for a very exciting research tool. A few of these studies have been conducted with respect to health and stress.

A recent study by Harari et al, an experiment conducted to study the behaviors patterns of activity and sociability over the course of a term. Data was collected via the StudentLife app of 48 healthy subjects. This was a preliminary study for this smartphone sensing application to measure the natural activity and sociability of students. The study focussed on these two behaviors, to identify when in term m-Health interventions might be most effective for students. After data analysis, it was discussed that individual sociodemographic predicted differences in behavioral patterns. It was also determined that, from the baseline established at the beginning of the study, the activity and sociability of the students went decreased before the midterm. After the midterm, it occurred that the activity levels stayed the same, but a big increase in sociability was detected [39].

At Dartmouth College, a study under the guidance of Wang was conducted to study the day-to-day and week-to-week stress and strain perceived by students. This was done by using the student life continuous data sensing application. It assessed the impact of workload on activity, mood, sleep, mental health, academic performance, and stress. During the data collection EMAs were sent to students, to identify subjective information about the student (e.g. mood and stress) The results showed that when the term progressed and the workload increased the students became increasingly stressed and sleep, activity and sociability was affected negatively [40].

Sano et al (2013), argued that it is possible to recognize certain stress parameters using wearable sensors (e.g. wrist sensor) and mobile phones. They aimed to find certain physiological and behavioral markers for stress. The study was performed with 18 subjects over the course of five days. At the beginning of the experiment, pre-surveys were filled out about the PSS, PSQI and the Big Five Inventory Personality Test. A wearable sensor was put on the non-dominant hand of the subject, this wearable collected three-axis accelerometer data and skin conductance. On mobile phones, calls, SMS, location, and screen on/off were monitored. Subsequently, every morning and evening certain surveys were filled out. After the experiment, another *post-experiment* survey had to be filled out. After data collection, data analysis was performed by implementing correlation analysis, this was done to find statistically proven significant aspects that are associated with stress. After that machine learning was applied to determine whether a participant was stress or not. A baseline accuracy was established at 87.5% percent using the pre-surveys. Using binary classification, the results showed an accuracy of 75%, this involved mobility, call, screen usage, and activity [41].

Although these studies show very exciting results, data collection through mobile data sensing is very sensitive. Personal information is not always protected, such as location, phone numbers, and certain habits. As a research tool, privacy and security of the personal data need to be guaranteed. A new platform that allows for continuous data sensing and user anonymity is called the AWARE platform. This platform will be used in this master thesis.

2.4 Assessing mobile data with the AWARE platform

Nowadays, mobile phones contain a lot of personal information and have the potential to sense the environment of its user, based on the sensors (i.e. luminance, accelerometer) present in mobile devices. The personal content of mobile devices makes it appealing for research. Nevertheless, there is a shortage of reusable and open software to build context-aware applications for mobile devices. This section will be used to describe the AWARE platform in terms of infrastructure, privacy, and security.

AWARE was created as an open network for context-aware computing research on mobile devices. AWARE allows you to collect data which could be used to find relationships and correlations between the data. This can include the individual and social behavior of the end-user. An AWARE application on a mobile phone is able to measure and collect various types of data using sensors (readily available in an Android phone) and plugins (e.g. more advanced sensors), such as hardware (i.e. accelerometer), software sensors (i.e. logging of calls and messages) and human-based data (i.e. questionnaires – through the experience sampling method(ESM)) [56]. The sensors and plugins serve the purpose of observing and monitoring collected data, the application makes it possible to find relations between different sensors of context information, such as social and individual behavior. Additionally, AWARE is a very special application as it ensures complete anonymity of the end-users, by encrypting every information transfer route and giving ids to the mobile device and specially encrypted ids to source information of calls and messages. This data collection can be performed as a result of users carrying their smartphone [32].

2.5 Willingness to participate in mobile data sensing studies

Before starting a mobile data sensing study, it is important to investigate the willingness of people to participate in such experiments. The reason for this that with mobile data sensing experiments data of the smartphones of participants is collected which is private information. The studies have been conducted to assess the willingness.

A study by Wenz et al, investigated the willingness of the general population to participate in miscellaneous data collection tasks using their mobile devices. 1660 members of the Understanding Society Innovation Panel filled in a survey including topics such as tracking apps, online questionnaires, how often a participant used a smartphone or a tablet. The results suggest that subjects were more willing to participate data collection that happened actively. Moreover, they were less willing to aid in tasks that require downloading and installing an application. In addition, it was found that the willingness was different between tablets and smartphones and varied between the subjects. The respondents who use their devices less frequently and the participants that reported higher security concerns with regard to mobile data collection were less willing to participate in such tasks [58].

Bouida et al. performed a study in 2016 to evaluate the willingness of the Tunisian population to participate in medical research. In addition, they tried to discover the factors that might influence the consent of the population. For this experiment three university hospitals used to recruit patients, healthy volunteers, and health care specialists to fill in a survey. This questionnaire was used to investigate factors that affected the consent for participation in medical research. It was concluded from this research that most young people or participants with a history of blood donation were willing to participate in medical research. This consent was lower for trials that dealt with life-threating diseases [59].

A study by Harari et al. investigated whether participants would be willing to give consent for the participation in studies that would track behavior. A group of students (N=1516) filled out a self-report survey to determine their willingness to participate in such a research. Of this 96% would be willing to take part in such experiments, of this percentage there was a large difference in willingness depending on the intrusiveness of the data. 54% would not mind receiving an EMA once or twice during the day, and 46% would be willing to provide access to sensors on smartphones. For access to social media accounts, 47% was willing to provide information and just 33% of the participants were willing to share web browser history.

The research states that concerns must be balanced against the merits of the experiment to encourage participants in smartphone sensing studies. The data from the survey also suggested that motivation might be able to serve as the motivation of participants to join such a study. Students reported different motivations for joining a self-tracking such as improving academic performance (80%), improving time management (63%), to understand when they are most productive (61%), keep track of stress and where the stress comes from (58%), etcetera. The survey suggests that it is possible to recruit participants for mobile data tracking, however further research is needed to determine the factors of these concerns, and what can be done about these concerns [66].

To the extent of knowledge, these are the studies with respect willingness to participate in mobile data sensing studies and healthcare studies. As not much research has been conducted on this topic it was decided to get an interview with a specialist in social sciences to obtain information from another perspective. What follows is an interview conducted with a domain expert to get information on how to investigate the willingness of students to participate in mobile data sensing studies.

2.6 Domain expert interview

On Wednesday, October 4th, Dr. M.L. Noordzij was interviewed. Dr. Noordzij is an associate professor at the Faculty of Behavioural, Management and Social Sciences (BMS). He was interviewed to get more information on student recruitment, the willingness of students to participate in seven-week mobile data sensing experiments with the inclusion of experience sampling (ESM). ESM according to the Catholic University of Leuven, is a method in which participants are asked to fill out questionnaires on multiple occasions throughout the day [61]. In addition, we talked about focus groups.

ESM was discussed first, Dr. Noordzij explained that ESM is often experienced as invasive. Most experiments in the BMS group that involve ESM have a duration of one to two weeks. It might be more motivating for students to participate in such experiments when students are rewarded with money for answering as much ESMs as possible. In addition, it is important to motivate the students with coaching. A possibility to keep the participants motivated in such experiments is by creating a WhatsApp group to keep students stimulated. Lastly, fixed times are more appreciated for ESM as it allows students/participants to create their schedule around the ESM times.

We talked about recruitment using a presentation, which he explained might be too little, he suggested to push and motivate students more often to participate in such studies. Moreover, a discussion about focus groups was held. Dr. Noordzij suggested performing one on one semi-structured interviews. This would be more effective to reach the objective of this experiment, as my experience with focus groups is very little. It seems that conducting a focus group can be very difficult when no experience is present. Therefore, the suggestion was presented by him to conduct individual interviews with eight to ten students for a duration of 20 minutes. When conducting interviews, ease the participants into it for the first couple of minute by asking some questions related to demographics. Lastly, it might make it more appealing for the students to participate in interviews when compensated for their time.

It was discussed that in psychiatric healthcare research the participant in mobile data sensing experiments would have some control over the data that is being measured during such an experiment. For example, the participant can choose different possibilities from a list or choose which parts of the day would be monitored. This might be a good idea to try out for a mobile data sensing experiment, especially in the case of a seven-week experiment.

Experiment 1: TECHNICAL AND USABILITY ASPECTS IN AN ANONYMOUS OBSERVATIONAL STUDY

This chapter describes the first experiment which aims to reach the first objective, which is to investigate the technical and usability aspects of the mobile data sensing platform. To be able to investigate this, the *mobile stress study* will be introduced as in input to investigate this objective. It was previously mentioned that to achieve this objective, three sub-objectives were investigated. The sub-objectives to be accomplished and why they are important are as follows:

O11: To determine parameters relevant to stress and to translate these parameters to sensors
In order to investigate stress through mobile data sensing it is important to find parameters that relevant to stress. In addition, these parameters should allow for stress monitoring though mobile data sensing.

O12: To investigate the jitter in sensor data
It is important to know that there are no large gaps in the collected data and the collection of data is consistent. This is an important aspect of data quality. Therefore, it will be investigated if the data is collected consistently (i.e. if jitter is present in the collected data).

O13: To investigate the response of the participants with respect to the questions sent via the Paco application
Literature reveals that information with respect to stress can be obtained from asking students questions. These are questions that cannot be obtained by mobile data sensing. This is with regard to for example, how someone's mood is or how someone is feeling at a certain moment. These questions will be asked using another application called Paco, which is a survey app.
The response is investigated in terms of response time to ascertain how fact students react to the questions. In addition, it will be investigated what period of the day most questions are answered. It is helpful to know at which time to ask questions and get the most responses possible. Lastly, the response is examined in terms of answered questions frequency. Investigating this is interesting to see if participants actually answer the questions. If more questions are answered the data collection is more complete and a better analysis will be possible.

O14: To investigate the usability of the Paco app, the AWARE app, and how the experiment was experienced by the subjects.
The usability of the mobile data sensing study should be investigated to discover how the participants have experienced the experiment. This should be investigated in terms of positive aspects and in aspects where improvement might be needed. Examining these aspects aids in proper execution of future mobile data sensing studies.

This chapter investigates these sub-objectives, after the results will be described, the results will be discussed and the sub-objectives will be assessed in the discussion.

3.1 Subjects

Students (i.e. male/female, ratio unknown) from the University of Twente owning Android smartphone, aged 18 years or older were recruited for this experiment. The recruitment for this experiment was realized by presenting a short presentation about the *mobile data sensing experiment study*. This presentation was held during two different lectures. One was held before first-year BME students and the second was held in front of first-year master students (primarily

BME students). After the presentation the students were given a web page (https://mterietstap.wixsite.com/awarepilotstudy) containing all the information given during the presentation, also containing the consent form to be signed in order for the participant to join the *mobile data sensing study*. The consent form needed to be signed to guarantee the anonymity of the participants, in addition, the subjects did need to provide their Gmail addresses to keep them informed about any changes that might have occurred during the experiment and to login to the Paco application. Gmail is necessary to be able to login to the Paco app.

3.2 Method

This section discusses the design of the study will be explained. First, the AWARE framework and the Paco survey app will be explained. Next, it will be explained how AWARE and Paco are involved in this experiment and how these apps are used to measure parameters and data relevant to stress. Lastly, a description will be given on the execution of the experiment.

3.2.1 AWARE platform

It was mentioned in chapter 2 that the AWARE platform is used to collect data on smartphone usage via mobile data sensing. This section provides in-depth information about AWARE.

AWARE's infrastructure

In Figure 3 an overview of the architecture is shown. Starting at the client, the AWARE application contains various sensors and plugins. The sensors and plugins collect data and the data is temporarily stored in an SQLite database, which is available on the participant's mobile device. When the mobile phone is connected to Wi-Fi, the data is sent as a JSON (JavaScript Object Natation) over MQTT-SSL, synchronizing the data. This allows for an encrypted and secure way to transfer the data to the MySQL database on the UT server. Going in the opposite direction starting at the server. On the dashboard, a new study can be designed selecting the plugins and sensors that will be used for data collection on smartphones. How such a study was created, see Appendix 1. When a participant joins a certain study, the specified sensors and plugins will be used for data collection on the smartphone of the client. The visualization of the AWARE app is illustrated in Appendix 1.

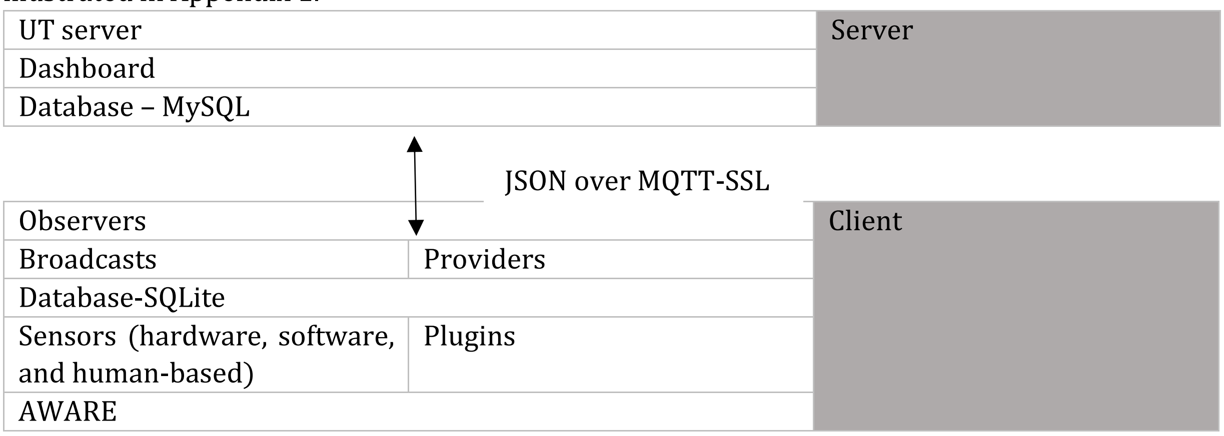


Figure 3: Overview infrastructure AWARE

Privacy

Privacy of the user is very important, this is why AWARE assigns a unique ID to each device and to all the contacts within the phone or unknown numbers. Additionally, no content (i.e. conversations, messages or any other content of applications) will be recorded, but rather if a conversation has taken place and its duration). These elements will ensure no personal information of the user will be known by the principal investigator.

It is important to guarantee the privacy of the data. A big merit of the AWARE framework is that in the same application data can be measured and questionnaires can be sent. This allows AWARE to assign a unique ID to each device, without requiring any information about the identity of the participant.

Security

The database is only accessible by the principal investigator, when a new study is created, special and unique credentials (i.e. Username and password) will be provided. This makes it a secure study on the UT server. The MQTT-SSL connection (SSL stands for Secure Sockets Layer, which is a cryptographic system protocol to send private data and documents over the internet [8], MQTT is an abbreviation of message queue telemetry transport which is used for sensor communication to remote locations (i.e. servers) is secure and the SQLite database is secured as it is part of the participants mobile device. To conclude, privacy and security will be available to ensure the data and the participants.

Anonymity

Starting the study; the subject will be asked to install the AWARE framework application. In this application, no registration is needed. Subsequently, a study will be activated through a QR-code and the device will then be connected to the study on the dashboard (i.e. connected to the UT server). The device has a unique visible ID on the dashboard, so it cannot be seen which smartphone belongs to which subject. Next, the subject will be requested to install additional sensors and plugins, showing notifications to gain excess to the conversation and applications etc. The content of these things will not be measured, only that a conversation takes place and which applications have been used (see Table 2).
During the study, all the measured parameters do not measure content, but rather that an event occurs. For example, if a user sends a text, it will be detected that a text has been sent instead of the content of that text. At the end of the study, the study can be stopped via the AWARE application by pressing the button: "quit study". The device will then be disconnected from the UT server.

The data will only be used for observation and no interventions will take place during this study. Additionally, the study is not completely anonymous as the name and e-mail address will be asked on the consent form. The e-mail address is necessary to come in contact with the participants if necessary. Next to the previously mentioned unique ID assigned to the smartphone of the participant. The contact information on the smartphone, such as phone numbers will have a unique ID as well. Every phone number gets a unique ID (encrypted). Lastly, the data measured on the smartphone of the participant cannot be seen by the participant.

3.2.2 Paco (survey app)

In this experiment, Paco will be used as a survey app that asks questions via smartphones throughout the length of the experiment. The specifics of the questions will be mentioned in another section. This section will explain the Paco app.
Paco is the application that will be used for the experience sampling. Paco is an open source platform which consists of a webpage (www.pacoapp.com) and a mobile app. via a special account, the principal investigator has made a study, which consists of the survey that will serve as the experience sampling.
The application requires participants to use a Google account to log in. Paco claims that having a google authenticity handle provides security. Fortunately, Paco allows the data to be downloaded anonymously.

It is important here to understand that a unique ID will be provided to every participant, just as for AWARE. These two IDs will be linked to each other with the help of a question send by Paco at the beginning of the experiment. This allows the AWARE data and the Paco data to be combined to the corresponding participant.

The server online shows a page with participants who joint the experiment (e-mail), but the retrieved data are represented by a unique ID that has been provided to each subject. The principal investigator is the only one who can see the e-mail addresses, the downloaded data will be anonymized. This is done by downloading the data anonymously from the Paco server as this is a possible option for Paco. The server used is the Paco server, this can only be accessed with special credentials that only the principal investigator has. Next, the data will be downloaded anonymously, the name of the participant will not be used, this is not of interest to us and it protects the participant's anonymity [62].

When the application is installed on the phone, there are a couple of things important. As stated above a login is needed. When signed in, there is a part were invitations to join an experiment are shown and also a part where can be seen in which experiments a subject is participating in, see Figure 4 and 5.

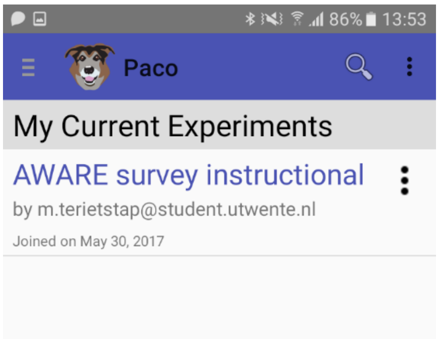


Figure 4: Paco - my current experiment page (snapshot)

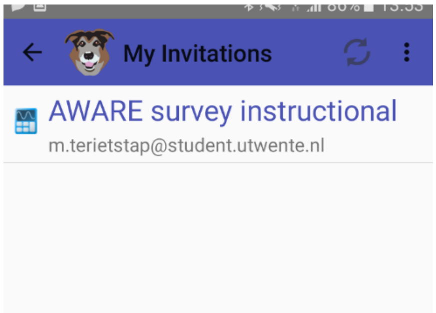


Figure 5: Paco - My invitations page (snapshot)

3.2.3 Suitable sensors and collected data (AWARE platform)

This section is divided into two parts. The first part describes what kind of sensors were used in the experiment relevant to stress, why they were used and an elaborate explanation of how the sensors work. The second part describes the same sensors and plugins in a more specific way with more technical details such as the sampling period.

3.2.3.1 Suitable sensors and plugins

Literature mentioned above shows that the following behavioral parameters are associated with stress and can be measured using a mobile data sensing experiment:

- Physical activity, as increased physical activity is associated with a reduction of stress;
- Social support, it is well known that interacting with other human beings can reduce stress levels;
- Subjective information, such as mood, feeling stressed, workload, etc. This is information that cannot be measured by parameters as it is human-based data, this is psychological data that can only be obtained from questions. These questions will be asked using ESM, which was done with Paco as mentioned in section 3.2.2.

Before listing all the parameters used for the study, it should be mentioned that all the devices involved in any way to this project (i.e. the device users, the devices called or messaged, and the devices in the surroundings) are all anonymized by assigning a unique encrypted ID to each of these devices. The parameters mentioned above will be measured using the next sensors and plugins, after the next section an overview is provided in Table 1 [35].

Activity recognition sensor (physical activity)
This plugin is able to detect what activity someone is performing, using an API from Google Location. This plugin can measure if someone is: still, walking, running, biking, or in a vehicle. It captures the activities in timestamps of milliseconds (ms), per device and it will show the activity name.
As social support is used in this study, mobile data sensing through social phone usage of students is performed. This is used as a measure of social interaction, for that different sensors and plugins will be used.

Application sensor
This sensor is able to log the use of notifications and applications on a smartphone. It tracks the applications running in the background and also tracks when the smartphone users switch between applications. The sensor provides the database with two different types of data tables. It uses a *foreground* table that shows a timestamp in milliseconds and which application is used at that time. In addition, it also captures the notifications received from the smartphone and detects when crashes of applications take place.

Communication sensor (social support)
It registers incoming and outgoing calls and messages. It should be noted here that the sensor anonymises the personal information such as the message contact or phone numbers. AWARE uses a special encryption (SHA-1 encryption) that assigns a unique ID will be given to the source. Incoming and outgoing calls and messages will be assigned a number, to mask content. For calls this means incoming calls are noted as 1 for incoming calls, 2 for outgoing calls and 3 for missed calls, subsequently messages are represented as 1 for received messages and 2 for sent messages. All these data are timestamped using milliseconds.

Device usage
A minimalistic plugin that detects when a phone is used or not used. It captures these sessions in periods of time, with a unit of milliseconds.

Bluetooth (measure of isolation)
A Bluetooth sensor is able to detect and log Bluetooth-enabled devices in its surrounding. If no surrounding devices are detected it can be assumed, that the owner of the device is alone. This is seen as isolation, it could provide information on the sociability or stress levels of the person. The specific sensor measures its surroundings with intervals of 60 seconds. Additionally, detected devices are assigned UUID (universally unique identifier, which is a 128-bit number). The timestamp of detection will also be provided for this sensor.

Table 1: Overview main elements of data collection

Sensor or plugin	Category	Data type
Activity recognition plugin	Physical activity	Kind of activity per unit of time
Application sensor	Social interaction	Apps used per unit of time
Communication sensor	Social interaction	Occurring texts and calls per unit of time
Device usage plugin	Social interaction	Phone use per unit of time
Bluetooth sensor	Measure of isolation	Detects surrounding devices per unit of time

3.2.4 Data collection

The different sensors and plugins mentioned above are presented in Table 2, which provides an overview of all information including the sampling period.

Table 2: Overview parameters and sensors

Parameter	Plugin or Sensor	Units of measurement	Sampling period (s)
Physical activity	Activity recognition (plugin)	Activity name: unknown, tilting, on_foot, in_vehicle, on_bycicle, running, walking Over a period of time (milliseconds)	60
Social support	Application sensor	Logs application running on foreground, notifications received and application history per time period (milliseconds) – <u>Note: All applications will be monitored</u> The next things will be logged: - Name of application (neither the content of the app is measured nor logged) - Name of notification neither the content of the notification is measured nor logged)	30
	Device usage (plugin)	Amount of time the device on (milliseconds) Amount of time the device is not being used (milliseconds)	
	Communication sensor	Collects calls and messages incoming and outgoing. - The device involved in the experiment, the contacts, and unknown numbers are assigned a unique encrypted ID. This ensures that the contact/subject information is unknown by the principal investigator. Data collected for phone calls Call_type (1 – incoming, 2- outgoing, 3- missed) (not the content of the calls is collected) Time period: length of calls (milliseconds) Data collected for messages Message_type (1-received, 2 – sent) (no the content of the messages is collected) In addition, the calls and messages being monitored, the correlated encrypted contact ID is monitored as well.	
Measure of social isolation	Bluetooth sensor – a measure of isolation	This sensor detects surrounding devices that are Bluetooth enabled. It assigns special ID's (encrypted) to the Bluetooth enabled devices (it does this at the time of occurrence).	60

3.2.5 Survey questions (Paco app)

As mentioned in section 3.2.3.1 subjective data (aka human-based data) were relevant to stress. Objective data result in questions as sensors cannot measure these types of data. For this, the Paco application was used to conduct experience sampling with relevant questions with respect to stress.

It was chosen to establish a baseline first, using the *Baseline Questionnaire* in Table 3.1. This baseline consists of entering the personal Device-ID of AWARE to connect the corresponding Paco data and AWARE data. Next, a question was implemented if students wanted to be informed if any patterns would have been recognized in their data. The last questions consist of the Global Assessment Recent Stress (GARS) scale (i.e. it assesses the current stress state of an individual), the specifics of this scale will be explained below in section Appendix 3. Filling this in before the experiment starts helped to establish a baseline.

Throughout the day experience sampling was conducted this was done five times a day, please see Table 3.2 for the content of the questions. The content of the questions was the same every day. Each time an ESM was sent to the smartphone it consisted of two questions. The first question always was: “Please rate how stressed you feel right now”, and the second question was related to factors that were associated with stress. The questions were based on findings from the literature. These questions could be answered with either yes/no or rated with a ten-point Likert-scale. The scale starts at zero and ends at nine, this was done because the GARS scale uses the same Likert-scale, which kept the rating consistent for the participants.

Every evening one ESM was sent to the participant, see Table 3.3 for the content. It was chosen to use statement one and statement eight from the GARS scale. This to keep up with the current stress levels of the participants.
The *final-survey* was conducted at the end of the experiment. This contains all of the GARS scale statements to obtain the currently experienced stress after the experiment. The content of the *final-survey* can be seen in Table 3.4.
Lastly, after the experiment was concluded a *post-experiment questionnaire* was sent to the participants. This to find out how the experiment was experienced by the participants. This questionnaire consisted of the System Usability Scale (SUS), which will be explained in section 3.2.6, and questions with respect to the content of the experiment. This *post-experiment questionnaire* can be seen in Appendix 2.

Table 3.1: Baseline Questionnaire

Baseline Questionnaire		9:00 DAY 1
Question/statement number	**Question or statement**	**Type of answer**
1.	Please enter your Device-ID. This can be found on the homepage of the AWARE-app.	Open answer
2.	If patterns of information are recognized in the data collected by the AWARE app on your smartphone, would you like to be informed?	Yes or no
3.	Please rate the next statement: Pressure related to work/job/school (whether self-imposed or not)	10 Likert scale 0 (None) – 9 (Extreme)
4.	Please rate the next statement: Pressure in interpersonal relations (family members and/or significant persons)	10 Likert scale 0 (None) – 9 (Extreme)
5.	Please rate the next statement: Pressure caused by sickness or injury (self, others, or both)	10 Likert scale 0 (None) – 9 (Extreme)
6.	Please rate the next statement: Pressure caused by financial issues	10 Likert scale 0 (None) – 9 (Extreme)
7.	Please rate the next statement: Pressure from unusual happenings (crime, natural disaster, accident, moving, etc.)	10 Likert scale 0 (None) – 9 (Extreme)
8.	Please rate the next statement: Estimate of overall level of pressure over the past week	10 Likert scale 0 (None) – 9 (Extreme)

Table 3.2: ESM between 10:00 - 18:00 DAY 1 & 09:00 - 18:00 DAY 2 - DAY 7

ESM (during the day)		10:00 – 18:00 DAY 1 09:00 – 18:00 DAY 2 – DAY 7
Question/statement number	**Question or statement**	**Type of answer**
1		
1.	From (0) to 9 (Extreme), please rate how stressed you feel right now	10 Likert scale 0 (None) – 9 (Extreme)
2.	From (0) to 9 (Extreme), how tired are you?	10 Likert scale 0 (None) – 9 (Extreme)
2		
1.	From 0 (None) to 9 (Extreme), please rate how stressed you feel right now	10 Likert scale 0 (None) – 9 (Extreme)
2.	From 0 (None) to 9 (Extreme), please rate how happy you feel.	10 Likert scale 0 (None) – 9 (Extreme)
3		
1.	From 0 (None) to 9 (Extreme), please rate how stressed you feel right now	10 Likert scale 0 (None) – 9 (Extreme)
2.	Have you met or are you going to meet up with friends or family today?	Yes or no
4		
1.	From 0 (None) to 9 (Extreme), please rate how stressed you feel right now	10 Likert scale 0 (None) – 9 (Extreme)
2.	Did you exercise today or are you going to exercise later in the day	Yes or no
5		
1.	From 0 (None) to 9 (Extreme), please rate how stressed you feel right now	10 Likert scale 0 (None) – 9 (Extreme)
2.	Have you eaten breakfast today?	Yes or no

Table 3.3: ESM 19:00 - 22:00 DAY 1 - DAY 7

ESM (in the evening)		19:00 – 22:00 DAY 1 – DAY 6
Question/statement number	**Question or Statement**	**Type of answer**
1.	Please rate the next statement: Pressure related to work/job/school (whether self-imposed or not)	10 Likert scale 0 (None) – 9 (Extreme)
2.	Please rate the next statement: Estimate of overall level of pressure over the past week	10 Likert scale 0 (None) – 9 (Extreme)

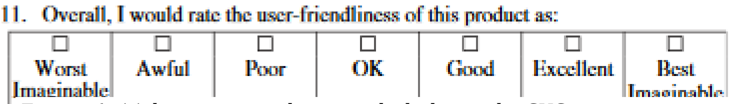
Table 3.4: Final-survey 20:00 DAY 7

Final-survey		20:00 DAY 7
Question/statement number	**Question or Statement**	**Type of answer**
1.	Please rate the next statement: Pressure related to work/job/school (whether self-imposed or not)	10 Likert scale 0 (None) – 9 (Extreme)
2.	Please rate the next statement: Pressure in interpersonal relations (family members and/or significant persons)	10 Likert scale 0 (None) – 9 (Extreme)
3.	Please rate the next statement: Pressure caused by changes in your relationships (death, birth, marriage, divorce etc.)	10 Likert scale 0 (None) – 9 (Extreme)
4.	Please rate the next statement: Pressure caused by sickness or injury (self, others, or both)	10 Likert scale 0 (None) – 9 (Extreme)
5.	Please rate the next statement: Pressure caused by financial issues	10 Likert scale 0 (None) – 9 (Extreme)
6.	Please rate the next statement: Pressure from unusual happenings (crime, natural disaster, accident, moving, etc.)	10 Likert scale 0 (None) – 9 (Extreme)
8.	Please rate the next statement: Pressure from change or lack of change in daily routine	10 Likert scale 0 (None) – 9 (Extreme)
8.	Please rate the next statement: Estimate of overall level of pressure over the past week	10 Likert scale 0 (None) – 9 (Extreme)

3.2.6 System Usability Scale (SUS)

The system usability scale (SUS) was developed by Brooke in 1996, which is designed as a quick survey to access the usability of a product or service easily [49]. There are quite a few other usability tests. However, the SUS has some real advantages. The first one is the fact that it can access wide ranges of different services and products (i.e. from simple software to more advanced software such as voice recognition). In addition, the way of scoring the survey is easy to understand from more advanced scientists to people who have very little experience with usability testing. Third, the survey is not protected by a brand name or trademark, making it very accessible to everyone. Lastly, the survey is quick and easy to use for everyone involved in the study or experiment. In Appendix 3 it can be seen that the SUS consists of ten statements using a five-point Likert-scale [49]. The total score can range from zero to 100, the total score implies that the higher the score the better the usability. SUS score can be calculated by adding up the contribution of each item first. The contribution of each item can range from zero to four. The odd items contribute by taking the position on the Likert-scale given and subtracting it by one. The even items are scored by taking five and subtracting it from the position given on the scale. Subsequently, sum all the contributions and multiply by two and a half, this gives the overall SUS value [49, 50].

In 2008 Bangor et al, conducted a study to help access the “goodness” of this test and to help practitioners with the interpretation of the SUS scores. First, the SUS was slightly altered by replacing “cumbersome” with “awkward” as not everyone seems to be familiar with the definition of cumbersome. This modification will be used in this experiment as well. An additional modification to the SUS was to add the 11th statement to access the overall user-friendly experience to make the results easier to interpret, see Figure 6 [50]. Adding the new statement was perceived as very positive and useful by 212 participants that completed the SUS with the additional 11th statement. This question has the intention to present a more qualitative answer, which can be combined with the SUS score to ensure a better explanation of the overall experience of the user interface's usability [50].


11. Overall, I would rate the user-friendliness of this product as:

☐ Worst Imaginable	☐ Awful	☐ Poor	☐ OK	☐ Good	☐ Excellent	☐ Best Imaginable

Figure 6: 11th statement that is included into the SUS

With the supplementation of these alterations, the question remains "What is an acceptable SUS score?" The research by Bangor has shown that a score that is at least above 70 is possible. Good products score in the upper 70s to high 80s. Products that score below the 70 mark should be considered for improvement. Scores below the mark of 50 should be considered unacceptable. Eventually, all the research performed in this study was combined in Figure 7 [50]. This figure helps to access whether a product or a service can be seen as acceptable or unacceptable. Within that range, it can also be seen "how acceptable" something is. It can be differentiated between *good*, *excellent*, and *best imaginable*. In conclusion, this usability scale will be used chapter 3 in addition to other questions that can be found in Appendix 3.

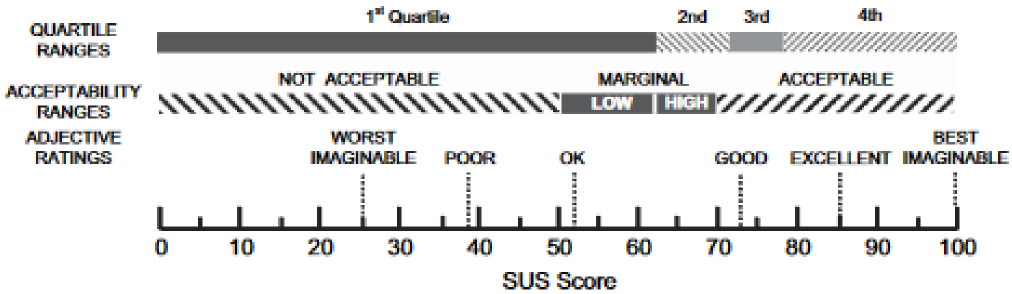


Figure 7: A comparison of mean System Usability Scale (SUS) scores by quartile, adjective ratings, and the acceptability of the overall SUS score

3.2.7 Mobile stress study

In this section, all the previously mentioned tools and materials were combined to describe how the mobile data sensing study was conducted. Here, the sensor data that measure parameters relevant to stress (through AWARE) and the questions relevant to stress (using Paco) ran parallel throughout the seven days of the experiment. On day eight, once more questions were asked through Paco while no data was collected through AWARE. This entire experiment is called the *mobile stress study* as previously mentioned in the introduction.

At the start of the experiment, the participants had to install the AWARE app and the Paco app using the instructional video on https://mterietstap.wixsite.com/awarepilotstudy/pilot-study-installation. After following the video the participants were enrolled in the mobile data collection from AWARE and Paco. The participants were asked to put their phones on vibrate, so no questions would be missed. Moreover, they were asked to keep their phone with them as much as possible and to make sure their phone would be charged throughout all of the experiment. The thing the participants had to do was answer the questions and let the phone collect the mobile usage data.

Day one of the experiment started with filling out a *baseline questionnaire*. Next, data was collected using AWARE and questions were asked during the day with Paco. These two

applications ran parallel throughout the entire experiment. The previously mentioned surveys and ESM were presented in a schedule when which survey was asked for the duration of the experiment, see Table 4.
Day two through five consisted of data collection by AWARE and the questions asked during the day by Paco. Day seven was the last day of data collection by AWARE and instead of the evening ESM, the *final survey* was sent to the phone.
Day eight was used to send a Post-experiment questionnaire to the smartphones of the participants. This questionnaire was used to assess how the mobile data sensing study was experienced by the students.
After day eight an e-mail was sent to all the participants to tell them how to quit the study on AWARE and Paco app.

Table 4: Schedule of ESM & questionnaires throughout the length of the mobile data sensing study

Subject	Days	Survey	Frequency	Time
	1	Baseline questionnaire ESM (during the day) ESM (in the evening)	1 5 1	09:00 10:00 – 18:00 19:00 – 22:00
	2	ESM (during the day) ESM (in the evening)	5 1	09:00 – 18:00 19:00 – 22:00
	3	ESM (during the day) ESM (in the evening)	5 1	09:00 – 18:00 19:00 – 22:00
	4	ESM (during the day) ESM (in the evening)	5 1	09:00 – 18:00 19:00 – 22:00
	5	ESM (during the day) ESM (in the evening)	5 1	09:00 – 18:00 19:00 – 22:00
	6	ESM (during the day) ESM (in the evening)	5 1	09:00 – 18:00 19:00 – 22:00
	7	ESM (during the day) Final-survey	5 1	09:00 – 18:00 20:00
	8	Post-experiment questionnaire	1	12:00

3.3 Results
This section presents the insights and analytical results from this experiment. The sensor data was analyzed to see whether the data was collected accurately (within the sampling period). The data from Paco was analyzed as well. It was chosen to not analyze the content of the answered questions, but rather to analyze the technical aspects of the data. In addition, feedback from the *post-experiment questionnaire* was analyzed to investigate the usability.

3.3.1 Subjects
The consent form to participate in the experiment was signed by six students. From the six students, five subjects accepted the invitation to join the Paco survey experiment. From these five participants, three subjects managed to install the AWARE application and join *the mobile stress study.*

3.3.2 Technical analysis of the survey data

The human-based data collected via Paco has been analyzed. The analysis was performed in three different ways. The first part was how long it took for the participants to answer question after the question was sent to the smartphone as a notification. The second component was to determine the number of questions answered and unanswered. The third component was to determine the distribution of the randomness in which the ESM questions were sent to the smartphones of the participants between 9:00 and 18:00 and at night between 19:00 and 22:00. This time range was divided into three-hour time windows, starting from 9:00 to 12:00, from 12:00 to 15:00, from 15:00 to 18:00, and from 19:00 to 22:00. These time ranges were chosen as the ESM during the day were sent between 09:00 to 18:00. Three-hour windows were chosen as it divides the morning, beginning of the afternoon and late afternoon in three equal parts. The window between 19:00 to 22:00 is simply the time window in which the evening ESM have been sent. The last component investigated, was in which time frame most questions were answered.

Figure 8 shows the period of time it took for participants to answer the questionnaire or ESM sent to their phones. The time period between the notification for each ESM/questionnaire and answering the question was calculated. The dotted black line focused on the y-axis is the average response time of five and a half minutes. The figure demonstrates that most of the received notifications were answered within the first six minutes. The black dotted line centered in the middle of the figure represents the snooze time, after which another notification was sent to the smartphone of the participant. Here it can be seen, that in most cases it was not necessary for the Paco app to push another notification to the smartphones of the participants.

Figure 9 shows a number of questions answered, a number of questions unanswered. Of the total amount of answered ESMs, 88% were answered before the second notification (reminder notification) was pushed to the smartphones of the participants.

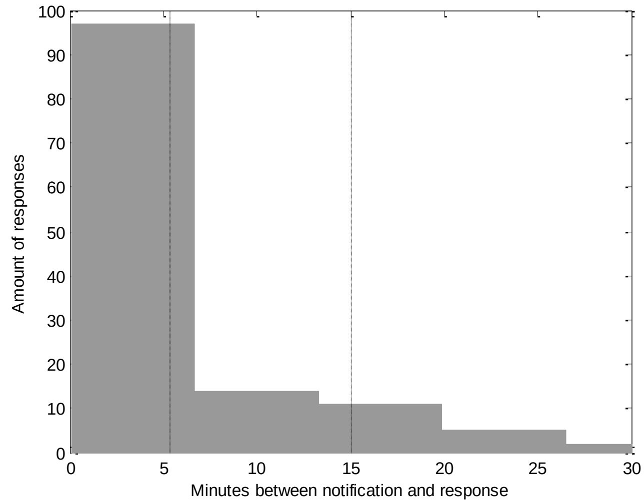


Figure 8: Elapsed time between notification and response

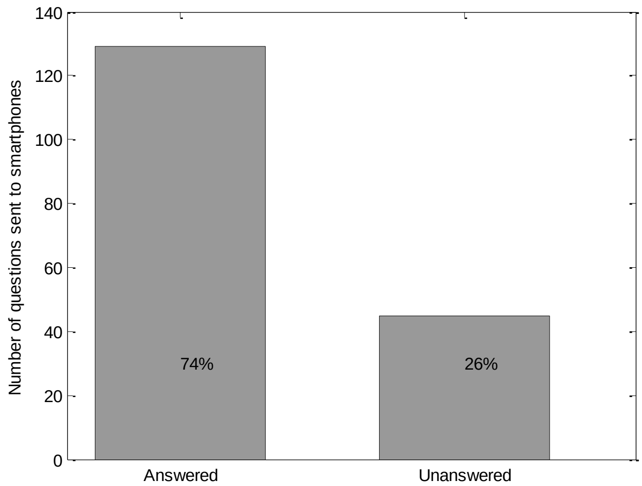


Figure 9: Notifications answered and unanswered

Figure 10 shows the amount of answered and unanswered questions sent to the smartphones of the participants during the experiment. The first three-time periods represent the ESMs during the day, and the last ESM is sent at night time. The percentages in the figure correspond to the number of ESMs answered within that period. The results show that the participants answered more questions between 12:00 and 18:00, than in the morning and at night.

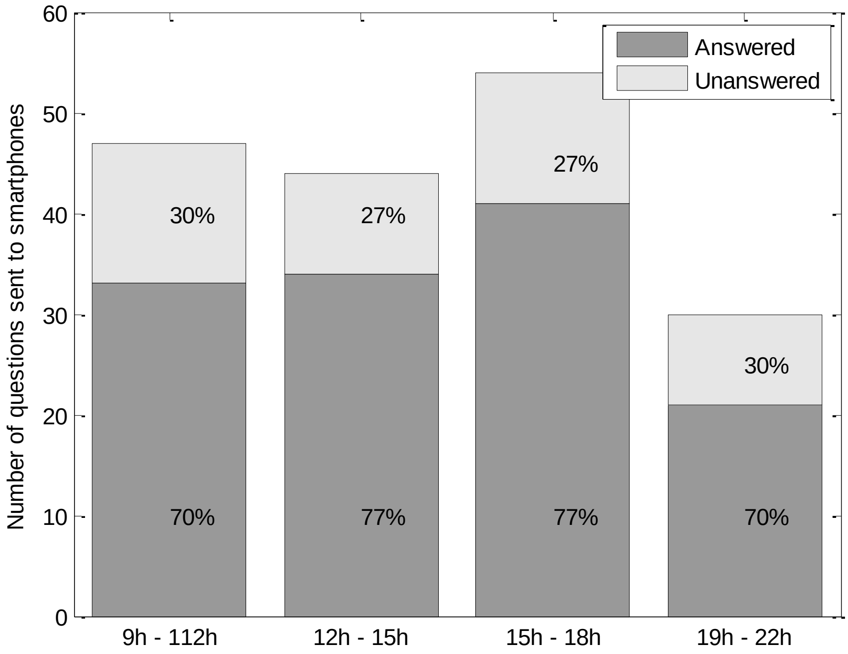


Figure 10: ESM notifications answered and unanswered within the time from of 9:00 to 12:00, 12:00 to 15:00, 15:00 to 18:00, and from 19:00 to 22:00

3.3.3 Technical evaluation of smartphone use

For the device usage, no sampling period could be set. Therefore, the data collection was tested in another way. The device usage was registered every time the screen of the smartphone was locked or unlocked and how long this event occurred in milliseconds. Therefore, it could be established if the milliseconds the device was locked or unlocked (in column 4 and 5) would correspond to the Unix timestamps in column 2 at which these events were registered, see Figure 11. An algorithm was designed to check if the milliseconds in row 2 column 5 were equal to the difference between the Unix time stamps of row 2 and row 1 in column 2. This was done for all the collected data of device usage.

2	3	4	5
'08-Jun-2017 16:00:03'	'513d2006-5e8c-4076-9d85-70ff328387c6'	0	0
'08-Jun-2017 16:01:01'	'513d2006-5e8c-4076-9d85-70ff328387c6'	0	58300
'08-Jun-2017 16:04:22'	'513d2006-5e8c-4076-9d85-70ff328387c6'	200837	0
'08-Jun-2017 16:04:26'	'513d2006-5e8c-4076-9d85-70ff328387c6'	0	4904

Figure 11: Snapshot device usage data

Table 6 shows the results of this algorithm. It can be seen here, that for each subject the start of data collection and end of the data collection is put in column 2 and 3(Unix time), respectively. Column 4 represents the window of the Unix time detection (i.e. the time between column 2 and 3). Column 5, represents the time window the device was registered on and off, here the milliseconds were converted to Unix time.

Table 6 shows that for each subject column 4 and 5 differ from each other. In an ideal case, this should not happen as it would have meant that on and off detection in milliseconds would be equal to the detected time window (column 4). However, column 5 has for both subjects a couple minutes of a difference with column 4.

To see what the cause of this difference was an analysis was performed. It was shown that the difference of one-time window when comparing the Unix time and the milliseconds when the device was either on or off could differ often from 10s to 100s of milliseconds. This might suggest round up mistakes. This algorithm showed some abnormalities, this meant that the difference in milliseconds sometimes was in the millions. Further investigation showed that these abnormalities were caused by turning off the phone completely. For this reason, these peaks were removed from the data collection. The data difference that can be seen in Table 6 is therefore only caused by the roundup difference.

Three subjects participated in this experiment, from Table 5 it can be seen that only the device usage data from two participants was used. It was decided to remove the third participant from the *device usage* dataset, as the data set was inconclusive. This means that there was so much data present that the participant should have been on his or her phone 24/7 without getting any sleep. This did not seem to be a realistic image.

Table 5: Collection overview dates of device usage per participant

Participants	Start data collection	End data collection	Time range (dd HH: MM: SS)	Calculated device on/off (dd HH: MM: SS)
Subject 1	10 – Jun – 2017 09:53:14	18 – Jun – 2017 20:36:00	08 10:42:46	08 10:43:49
Subject 2	08 – Jun – 2017 19:09:08	19 – Jun – 2017 07:30:29	10 12:20:21	10 12:27:27

3.3.4 Technical examination of physical activity sensor data

On the AWARE dashboard, the sampling period for physical activity measurement was set to every 60 seconds. To determine whether jitter was detected in the physical activity sensor data, all the time windows between two detection times were checked. An algorithm was designed to check whether the difference between two times was less, more or equal to the sampling period, 0 to 30 seconds delay, 30 seconds to one-minute delay, one minute to four minutes delay, or over four minutes delay. These ranges were chosen as the first range was the sampling period. The next two delays were very close to the sampling period and it was therefore chosen to make these periods smaller. The one to four minute delay was chosen as a larger window of time as these detections were less than the previously mentioned delays. After four minutes of delay very little was detected in comparison to the other delay detection windows. Therefore, it was decided to make this separate from the other delay time windows.

In Figure 12 it can be seen that each time stamp is connected to an activity, for example *still* or *tilting.* It was decided to see whether the difference between two timestamps (column 1) was connected to a specific activity. Two activities are involved with each timestamp difference, this time represents the delay in detection or detection within sampling period as mentioned previously. Therefore, it was analyzed whether these two activities were either consecutive activities (i.e. row one and two) or if the two activities were alternating activities (i.e. row three and four).

'04-Jul-2017 19:29:27'	'a78a697a-b352-45fa-b11f-f66ca595cd5a'	'still'
'04-Jul-2017 19:32:28'	'a78a697a-b352-45fa-b11f-f66ca595cd5a'	'still'
'04-Jul-2017 19:32:28'	'a78a697a-b352-45fa-b11f-f66ca595cd5a'	'tilting'
'04-Jul-2017 19:32:50'	'a78a697a-b352-45fa-b11f-f66ca595cd5a'	'still'

Figure 12: Physical activity detection example (snapshot)

Figure 13 shows the jitter in the sensor data. This was done to determine if the detection was performed according to the fixed sampling period. The first bar shows the detection within the fixed sampling period (no jitter), the second bar shows the detection between 0 and 30 seconds after the fixed sampling period has passed, etcetera. It can be seen here that 56,1% of the physical activity has been detected after a 0 to 30-second delay, and 35,9% has been detected within the fixed sampling period.

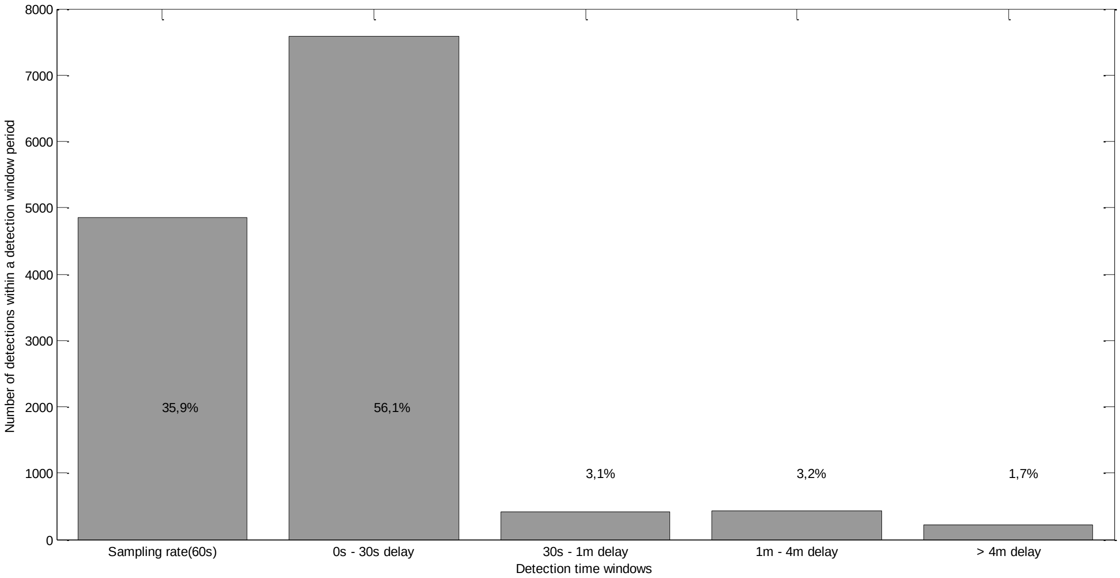


Figure 13: Overview: Detection of physical activity within and outside of the sampling period

For Figure 14 it was chosen to design an algorithm that would be able to find the number of detections per activity and per time window (i.e. either sampling period or a delay). In addition, for each activity, it was detected whether two successive activities where either alternating activities or the consecutive activities. Figure 14 shows alternating (i.e. two different activities) and consecutive (i.e. the same activities) that have been detected in successive order. No detection of the activity *walking* had occurred. In contrast to, the logging of the activity *on_foot* which has occurred in alternating and consecutive activities for each detection time window. In addition, the activity *tilting* has only be detected in alternating activity pairs. The idea behind this was to observe whether certain forms of activities occurred more or less within the sampling period or beyond the detection frequency. Figure 14 shows that the activity *still* has been logged most within and outside of the sampling period in comparison to the other logged activities. Beyond a delay of four minutes, the number of consecutive *still* activity detections has increased to 66,2% of the total number *still* activity detections beyond a four-minute delay. The observation for the *tilting* activity as only being logged as an alternating detection in Figure 14. The activities *still, tilting*, and *unknown* have been detected most often.
In addition, Figure 14 shows that in the > 4-minute delay section only *still* was detected as a consecutive activity. All other activities have only been logged as alternating detections. Lastly, *on foot* activity has been logged most often as an alternating detection.

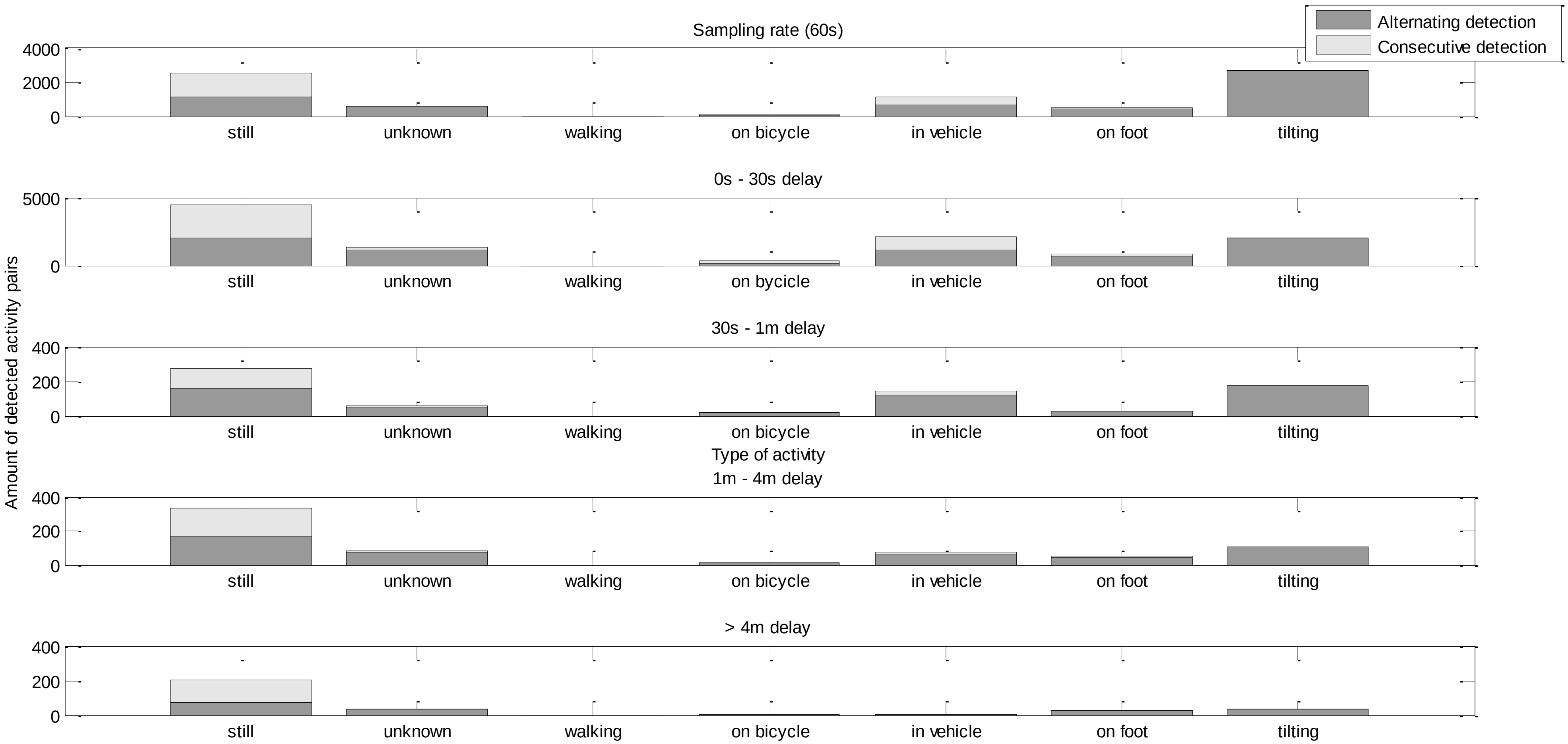


Figure 14: Amount of detected activity pairs (consecutive detection or alternating detection)

3.3.5 Technical investigation of application use

The investigation for app use was performed with the same analysis as the physical activity was analyzed with. Thus, to detect if there was jitter in the sensor data. However, the sampling period for app use was set to every 30 seconds. An algorithm was designed to check whether the difference between two times was less, more or equal to the sampling period, zero to 60 seconds of delay, 60 seconds to 90 seconds of delay, 90 seconds to three and a half minutes of delay, and beyond three and a half minutes of delay. These ranges were chosen as the first is the sampling period. The next two delays are very close to the sampling period and it was therefore chosen to make these periods smaller. The 90 seconds to three-and-a-half-minute delay was chosen as a larger window of time as these detections were less than the previously mentioned delays. After three and a half minutes of delay very little was detected in comparison to the other delay detection windows. Therefore, it was decided to make this separate from the other delay time windows. The results are presented in Figure 15. Figure 15, shows that 83,5% of all application use detections have been logged within the selected sampling period.

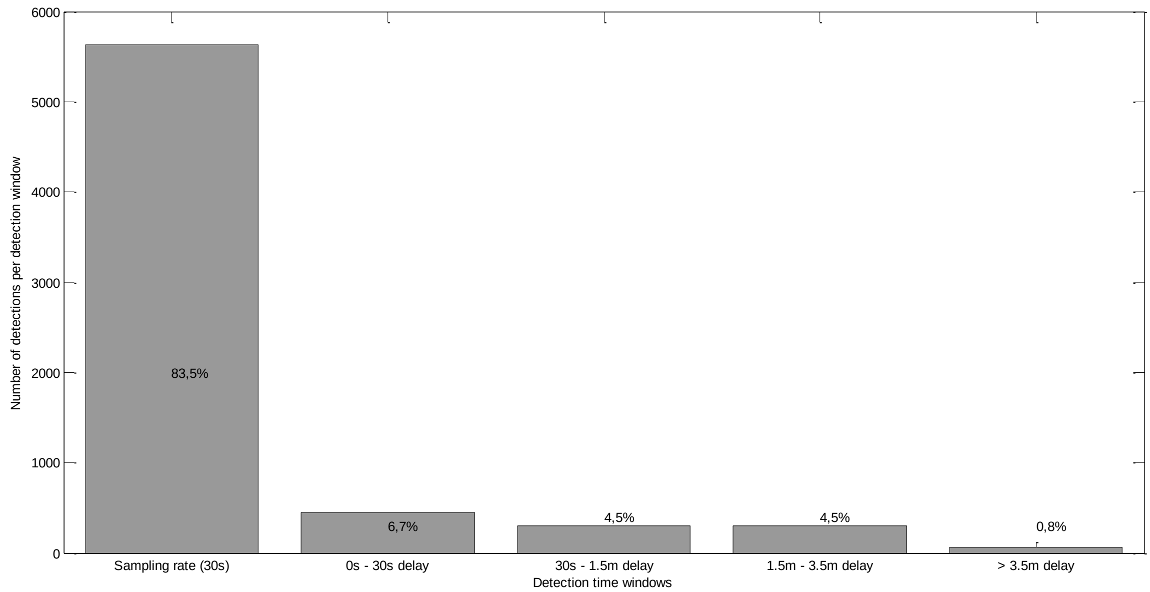


Figure 15: Overview application use detection per time window

3.3.6 Technical investigation of notifications, calls, messages, and Bluetooth (social interaction)

It is not possible to see whether any technical difficulties were present for the notifications that were sent to the smartphones of the participants. This is because for notifications data will only be collected when a notification has been received. Therefore, it was decided to check whether any notifications were received throughout the extent of the experiment, which was the case. However, the received notifications were collected for each participant. It was checked if the messages and the calls were registered during the data collection period. The calls and messages were only detected when an event happened. For each participant, calls and text messages have been logged.

The technical analysis of the Bluetooth sensor in the second experiment was the same analysis performed in the first experiment. To detect possible jitter in the sensor data. On the AWARE dashboard, the sampling period for Bluetooth measurement was set to every 60 seconds. To determine whether the sensor detected the data accurately (i.e. measured within the sampling period), all the time windows between two detection times were checked. An algorithm was designed to check whether the difference between two times was less, more or equal to the sampling period, 0 to 30 seconds delay, 30 seconds to one-minute delay, one minute to four

minutes delay, or over four minutes delay. These ranges were chosen as the first range was the sampling period. The next two delays were very close to the sampling period and it was therefore chosen to make these periods smaller. The one to four-minute delay was chosen as a larger window of time as these detections were less than the previously mentioned delays. After four minutes of delay very little was detected in comparison to the other delay detection windows. Therefore, it was decided to make this separate from the other delay time windows.
Figure 16 shows that most detection has been performed at the sampling period and the first 0 to 30-second detection window of delay. Together these have collected 85,8% total collected data. The results showed that the Bluetooth should have been activated manually, even though AWARE made it appear as if Bluetooth would be automatically activated.

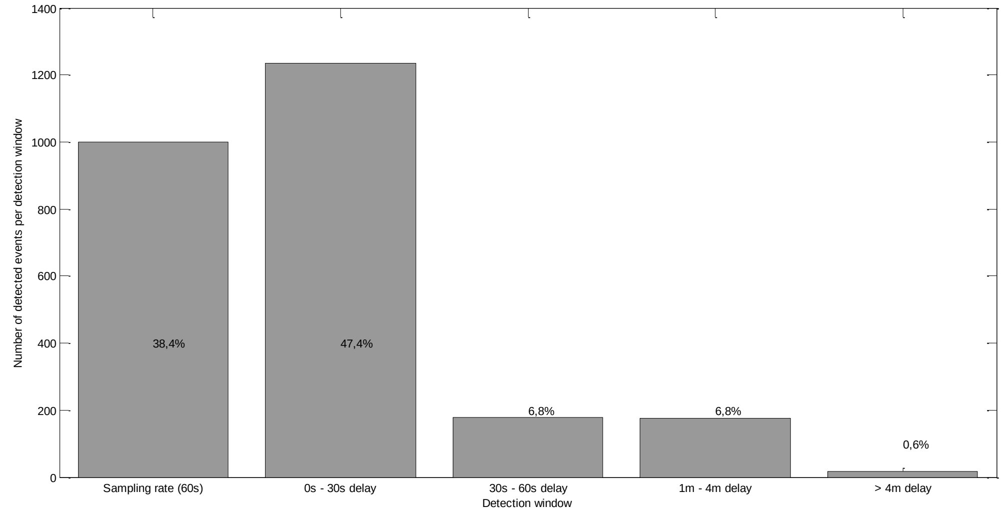


Figure 16: Overview Bluetooth events logged per detection window

3.3.7 Usability assessment post-experiment questionnaire

The data from the *post-experiment questionnaire* was analyzed and used as feedback to see how the system usability was perceived by the subjects. The first eleven statements of the survey were to determine the usability system according to the SUS survey as previously explained. The remaining questions and statements in this survey were used to establish the experience of the participants with regard to the *mobile stress study*.
The *post-experiment questionnaire* was answered by one subject, the answers from this survey are presented in Appendix 4. The first ten questions of this survey consisted of the System Usability Scale (SUS). The calculated score from the SUS survey was 70. According to Figure 7, a score of 70 can be interpreted as an acceptable study in terms of usability. In addition, the overall user-friendliness was scored as OK, which is equal to a score of a little over 50, according to Figure 7.

The content of the questions was not perceived very well. According to the participant, the content of the questions had little to no relationship with stress. The subject thought the content of the questions should include the next topics, *"I perceive stress when someone reminds me how much stuff I have to do"*, and " *When I do something with friends while I would be studying"*.
The participant would like to have the notifications available on the smartphone screen for longer (i.e. several hours longer). So, when the phone does not have to be on vibrate all the time and no questions would be missed.

3.4 Discussion

In what follows the results of the separately evaluated parts have been discussed.

3.4.1 Evaluation results of the survey data

The average response time was determined to be 5.30 minutes. This was a decent result as the participants had 30 minutes to answer a notification with a reminder at the 15-minute mark. 74% of all the questions sent to the smartphones were answered. From the answered questions, 88% were answered before the reminder notification around 15 minutes after the question had been sent to the smartphone. These are promising results, as it indicates that when participants take part in mobile data sensing studies they do respond to most questions and also very quick.
Most questions were answered between 12:00 and 18:00, this might suggest participants are more prone to answer questions in the afternoon or it might have to do with the segmentation of the detection periods. Next, the sensors and plugins have been discussed, Table 6 shows a summary of this discussion.

3.4.2 Evaluation results of smartphone use

Table 5 suggests that roundup deviations with respect to time occurred every time a detection was logged. Column 2 contained timestamps in Unix time and column 3 contained just milliseconds. These different measures of time result in roundup deviations. Therefore, for future research, only the Unix timestamps should be used. No gaps were found in the data collection, which indicates consistency of the data.

3.4.3 Evaluation results in physical activity sensor data

The results of Figure 13, showed that most activities were detected beyond the sampling period (35,9%) in the 0 to 30-second delay (56,1%). This suggests that there is just a small presence of jitter. This slight deviation from the sampling period might be a result of Android determining when an activity was made accessible to AWARE. It was discussed that in future research the sampling period would increase the sampling period to every 30 seconds. This was decided to be able to detect all changes in activities throughout the data collection process.

Figure 14 showed that no detection of *walking* had been logged. AWARE used Google Location API to capture activity [55]. For this reason, *walking* has been registered as an *on_foot* activity which is a characteristic of Google Location API. A sub-activity of *on_foot* is walking or running, therefore *walking* is detected as *on foot* [54]. The results suggested that tilting has only be detected in alternating activity pairs. This suggests that the phone was tilted/held by the subject just before alternating activity.
Striking was that *on foot* detection was always captured within an alternating detection pair. In combination with a large number of consecutive and alternating *still* detections, it might be suggested that the participants sat down a lot. Especially, the detections in the time window beyond a 4-minute delay contained a large number of consecutive still detections. This suggests that phones were not used and were put down on tables for long periods of time.

3.4.4 Evaluation results of application use

The app sensor works fine, as 85% of the data collection was performed within the sampling period. This means that a very small part of the sampling data contains jitter. Therefore, it might be suggested to leave the detection frequency the same. However, as there were a lot of detected events, as a result of that the sampling period was decreased to 90 seconds.

3.4.5 Evaluation results of notifications, calls, messages, and Bluetooth

Notifications were collected for every participant. As the notifications received by the subjects were collected it can be concluded that the notifications are collected well. Calls and text messages were collected by AWARE. It can however not be traced back if some calls or messages were not collected as the experiment was performed anonymously.
The Bluetooth sensor was not manually enabled, meaning that no devices were detected. In future experiments, the sensor should be enabled manually. The Bluetooth sensor collected the sampling data with just a small jitter. No large data gaps suggest pretty consistent data collection. Figure 16, showed that 38,4% of detected data was collected within the sampling rate. In addition, 47,4% was detected in the first 0 to 30-second delay. This seems to be very promising.

Table 6: Summary performance sensors and plugins

Sensor/plugin	Operation	Remarks
Device use plugin	Good	
Activity recognition sensor	Good	60 second sampling period becomes 30 second sampling period
Application sensor	Good	30 second sampling period becomes 90 second sampling period
Communication sensor	Good	
Bluetooth	Good	Bluetooth should be enabled manually

3.4.6 Evaluation results usability assessment post-experiment questionnaire

The feedback from the post-experiment questionnaire resulted in some interesting findings. The SUS score was 70, which was acceptable. The design of the mobile stress study might be improved for future studies to increase the usability and experience of the study.
According to the *post-experiment questionnaire*, the content should be changed as it was not experienced as relevant to stress. The questions, however, are relevant, as they are all backed up by research. This should be noted in the next stress modeling system study. In addition, more research should be done regarding the content of the questions to make the experience as relevant and as possible.

The survey suggested that there should be a possibility to answer the questions afterward, when or if a notification was missed. This can simply be done by designing a Paco experiment such that the notifications will be available for longer on the smartphones of the participants.

3.4.7 Study limitations

The results of experiment need more participants to be generalized to all students. However, some very interesting findings have been done. In future research, the Bluetooth should be enabled manually. The *post-experiment questionnaire* would have provided more significant results if it had been filled out by all participants of the study.

3.5 Conclusion

The final step of this chapter, the sub-objectives as defined at the beginning of the chapter are examined. As previously mentioned to fulfill the objectives of this thesis, the thesis is divided into three experiments each experiment has its own sub-objectives. In what follows the sub-objectives of experiment 1 will be discussed:

O11: To determine parameters relevant to stress and to translate these parameters to sensors

To understand the stress and the parameters that are relevant for stress detection, a literary study had to be conducted. Literary research has shown that there are various factors related to stress. However, current factors are mostly related towards self-report or physiological measurements. These factors cannot be used for sensor data, therefore, parameters related to stress-management were used to choose these parameters. These parameters were as follows:

- Physical activity, as increased physical activity is associated with a reduction of stress;
- Social support/interaction, it is well known that interacting with other human beings can reduce stress levels;
- Subjective information, such as mood, feeling stressed, workload, etc. This is information that cannot be measured by parameters as it is human-based data, this is psychological data that can only be obtained from questions.

Physical activity and social support/interaction can be translated to sensors and plugins of which data can be collected using AWARE. Meanwhile, subjective information can be obtained with surveys and ESM, was done using the Paco app.

Combining the parameters relevant to stress with the sensors and plugins available for AWARE was a process. It was chosen that data related to physical activity could be collected using the activity recognition plugin. Social support/interaction can be measured in various ways on smartphones. The sensors and plugins used for that were:

- Device usage plugin, to see how often a subject uses his or her smartphone;
- Communication sensor, to see how often a subject is being called, calls, sends a text, or is being sent a text;
- Application sensor, to see what apps the subject uses. This with a specific interest in the social media apps or communication apps such as WhatsApp.
- Bluetooth sensor, which can be used as a detection method for isolation analysis.

These were the sensors used to measure the parameters relevant to stress.

O12: To investigate the jitter in sensor data

The jitter in the sensor data of the physical activity, Bluetooth, and application sensor was found to be minimal. This because most data was collected within the sampling period or in the first-time window of delay (0 to 30 seconds). This suggests that there are no large inconsistencies in the data collection and implies that data quality is good.

This jitter in terms of sampling period cannot be analyzed for the device usage plugin; however, the data was collected properly as no data gaps were found. The same goes for the communication sensor, however, here it was registered when a call or a text message took place. Table 7 shows an overview of the sensors.

Table 7: Overview all sensors

Sensor/plugin	Performance	Remarks
Device usage plugin	Good	
Activity recognition sensor	Good	60 second sampling period becomes 30 second sampling period
Application sensor	Good	30 second sampling period becomes 90 second sampling period
Communication sensor	Good	
Bluetooth	Good	Bluetooth should be enabled manually

O13: To investigate the response of the participants with respect to the questions sent via the Paco application

The results showed that the average response time of the participants was 5.30 minutes, this is a very interesting result. As it suggests, when students are participating in an experiment such as the *mobile stress study*, that they respond quickly. In addition, 88% of the questions were answered before a reminder was sent to the smartphones at the 15-minute mark.

The answering rate amongst the subjects was pretty high. 74% of all send questions were answered.

The questions were evaluated in four equal windows of time. One representing the morning, one the beginning of the afternoon, one the end of the afternoon, and one representing the evening. It was found that most questions were answered in the afternoon. 77% of the questions send between 12:00 and 15:00 were answered and also 77% of the questions send between 15:00 and 18:00 were answered.

O14: To investigate the usability of the Paco app, the AWARE app, and how the experiment was experienced by the subjects.

A score of 70 was given to the usability of this *mobile stress study*, which translates to an acceptable study according to SUS. From the SUS score it can be translated hat the system was experienced to be easy in use, the system was found to be easy to learn.

A score of OK was given to the overall user-friendliness of the system according to SUS. This means that the study is OK but that some improvements might be suggested for future experiments. A couple of suggestions were given to improve future studies:

- To have to notifications send by Paco available for longer on the smartphone, so the smartphone does not have to be on vibrate constantly and no questions will be missed;
- The questions should be more relevant to stress. However, the questions are all related to scientific evidence. In future studies, this should be emphasized better.

Reflecting on the main objective, it is concluded that these technical and usability evaluations could be useful for future mobile data sensing studies. This experiment is only part one to obtain an answer to the first objective, therefore more results and conclusions will follow in the next chapters.

Experiment 2: TECHNICAL AND USABILITY ASPECTS IN A CONTROLLED STUDY

This chapter explains the second experiment, which aims to fulfill the first objective. This is to investigate the technical (and usability) aspects of the mobile data sensing platform. In this experiment, the *mobile stress study* will be introduced once. However, in this experiment, a few alterations have been applied. First, the Paco app has not been included, which means no experience sampling did not occur. Second, the mobile stress study has an experiment length of two days. Lastly, as the experiment is conducted with smartphones that are the property of the university, no communication sensor or applications sensor was used. In addition, the Wi-Fi sensor and the battery sensor have been included. It was previously mentioned that to reach this objective, three sub-objectives were investigated with reasons why, these are listed below:

O21: To assess the usability of the system.
This is important with respect to the installation process of AWARE. This makes it possible to assess the experience of the participants and the investigate potential improvements. Eventually, it will make the process of enrolling in an AWARE experiment for future research easier.

O22: To investigate the battery drainage caused by AWARE.
It is useful to know if participants might need to charge their phones more than once a day during a mobile data sensing experiment with AWARE to ensure sufficient battery levels for data collection. This to ensure that all data is collected during the experiment and no data will get lost.

O23: To investigate if there is jitter in the Wi-Fi sensor data.
Wi-Fi scans can be used to track location. It was decided to use location tracking as changes in location might influence differences in stress. Therefore, the jitter in the sensor data should be investigated to see whether all the data is collected.

In addition, the results will be described, the results will be discussed and the sub-objectives will be assessed in the conclusion.

4.1 Subjects

Five students (four male, one female) of the University of Twente were randomly recruited by asking them face-to-face. For this experiment, it was not necessary to sign a consent form, as the phones used were not owned by the participants but by the University and oral consent was given.

4.2 Method

This section discusses the design of the experiment which included: the data collection, materials used, and the execution of the experiment.

4.2.1 Data collection

The study made on the AWARE dashboard was supplemented with a few sensors and plugins. For this experiment, the plugins and sensors differ from the first experiment as smartphones used do not contain the apps used on regular phones. Moreover, no calls or text messages would be sent to these smartphones. Subsequently, a Wi-Fi sensor and a Battery sensor were included.
Being aware of the locations of a participant could provide insightful information when analyzing the other parameters. This sensor (via Wi-Fi sensor) estimates the current location of the user. It

detects the network name the smartphone is currently connected to, this allows for location triangulation. The battery sensor was included to see how fast the AWARE app drains the battery of the phone and the Wi-Fi sensor was used to investigate if there was jitter in the location tracking sensor data. Table 8 shows an overview of the plugins and sensors used in this experiment.

Table 8: Overview sensors/plugins used in experiment

Parameters	Plugins/sensors	Units of measurement	Sampling rate (s)
Location	Wi-Fi sensor	Logs mobile device's Wi-Fi sensor, current AP and surrounding Wi-Fi visible devices. - Ssid: connected access network name (at a certain time)	60
Social interaction	Plugin: Device Usage	Amount of time the device on (milliseconds) Amount of time the device is not being used (milliseconds)	
Physical activity	Plugin: Activity Recognition	This plugin automatically recognizes which activity is taking place. The activities it recognizes is walking, in the vehicle, on a bicycle, and running. Otherwise, it will be recognized as unknown. Activity name: unknown, tilting, on_foot, in_vehicle, on_bycicle, running, walking Over a period of time (milliseconds)	60
Measure of isolation	Bluetooth sensor	This sensor detects surrounding devices that are Bluetooth enabled. It assigns special ID's (encrypted) to the Bluetooth enabled devices (it does this at the time of occurrence).	60
Energy consumption of AWARE app	Battery sensor	Monitors battery information and power related occurrences. It collects the following data: - Current battery voltage	60

4.2.2 Materials

To perform the experiment, smartphones were used that were the property of the University of Twente. Table 9, gives an overview of the characteristics of the smartphones [57].

Table 9: Smartphone specifications

Models / **Specifications**	Google Pixel	LG G5	Archos 45 platinum	Samsung Galaxy J5 (2016)	Motorola Moto G3	Huawei P9	Oneplus 5T	Samsung Galaxy S7
Processor	1.6GHz quad-core	2.15Ghz quad-core	1.4GHz quad-core	1.2GHz quad-core	1.4 GHz quad-core	1.8GHz octa-core	2.45GHz quad-core	1.6GHz octa-core
RAM	4 GB	4 GB	512 MB	2 GB	2 GB	3 GB	6 GB	4 GB
OS	Android 7.1	Android 6.0.1	Android 4.4.2	Android 6.0.1	Android 5.1.1	Android 6.0	Android 7.1.1	Android 6.0
Battery (mAh)	2770	2800	1700	3100	2470	3000	3300	3000
Storage (GB)	32	32	4	16	8	32	64	32
Resolution (pixels)	1080x1920	1440x2560	480x854	720x1280	1280x720	1080x1920	1080x2160	1440x2560
Display (inch)	5.00	5.30	4.50	5.20	5.00	5.20	6.01	5.10
Camera (megapixel)	8 (front) 12.3 (rear)	8 (front) 16 (rear)	5 (rear)	5 (front) 13 (rear)	5 (font) 13 (rear)	8 (front) 12 (rear)	16 (front) 20 (rear)	5 (front) 12 (rear)

4.2.3 Execution of experiment 2

The duration of experiment 2 was two days in which the participants were given one or two Android smartphones each. Each student was individually asked and guided to watch the instructional video designed for the installation of the AWARE app and asked to install the app, on https://mterietstap.wixsite.com/awarepilotstudy/pilot-study-installation. Afterward, a face-to-face conversation was held by the principal investigator and the subject to discuss aspects of the installation process, instructional video, and the AWARE application. Students are asked to keep the smartphone(s) with them for the next two days. Afterward, the phones were taken back and the experiment was ended.

4.3 Results

This section presents the usability and analytical results of this experiment. The sensor data of the battery sensor, the Wi-Fi sensor, and the Bluetooth sensor were analyzed. In addition, the usability as assessed using face-to-face discussions.

4.3.1 Technical investigation battery sensor

Analysing the battery drainage was necessary to see how much energy AWARE consumes. For the analysis three of the smartphones were used as these provided the best data. Figure 17 shows that battery level on the y-axis and the time on the x-axis. This figure was an approximation of the battery level drainage (max battery level is 100). The graph shows that over the course of a day of detection and no charging of the smartphone. The phones were not used during the day, not all plugins were installed, and Paco was not installed on the smartphone. The graph shows that the phones remained with battery levels between 30% to 40% and the smartphones started off the experiment with a percentage of 75% to 85%. The results show that after a period of a 24-hour period the phones still had battery left. This means that AWARE app drains the battery approximately 40% to 60% in a 24-hour period.

The literature showed that the Paco application does not drain the battery a lot, therefore it was not tested [62]. However, during the *mobile stress study*, the Paco app is running parallel with the AWARE app and the phone will be used as well.

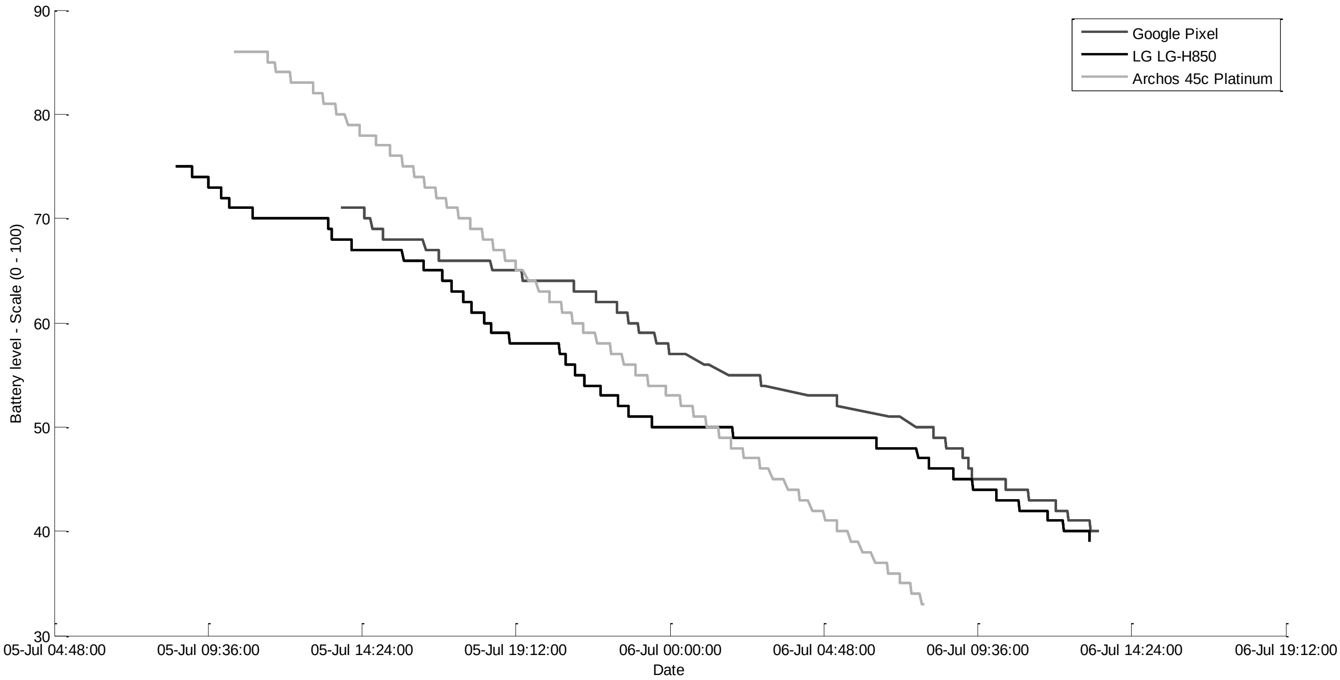


Figure 17: Battery drainage of three smartphones

4.3.2 Technical analysis of Wi-Fi sensor

On the AWARE dashboard, the sampling period for location measurement was set to every 60 seconds. To determine whether the sensor detected the data accurately (i.e. measured within the sampling period), all the time windows between two detection times were checked. An algorithm was designed to check whether the difference between two times was less, more or equal to the sampling period, 0 to 30 seconds delay, 30 seconds to one-minute delay, one minute to four minutes delay, or over four minutes delay. These ranges were chosen as the first range was the sampling period. The next two delays were very close to the sampling period and it was therefore chosen to make these periods smaller. The one to four-minute delay was chosen as a larger window of time as these detections were less than the previously mentioned delays. After four minutes of delay very little was detected in comparison to the other delay detection windows. Therefore, it was decided to make this separate from the other delay time windows. The results are shown in Figure 18. Figure 18 shows that most data detected within 0 to 30-second delay and within the sampling period. That is a total of 94,5% of all logged event detections. This means that the jitter for the Wi-Fi sensor data is minimal.

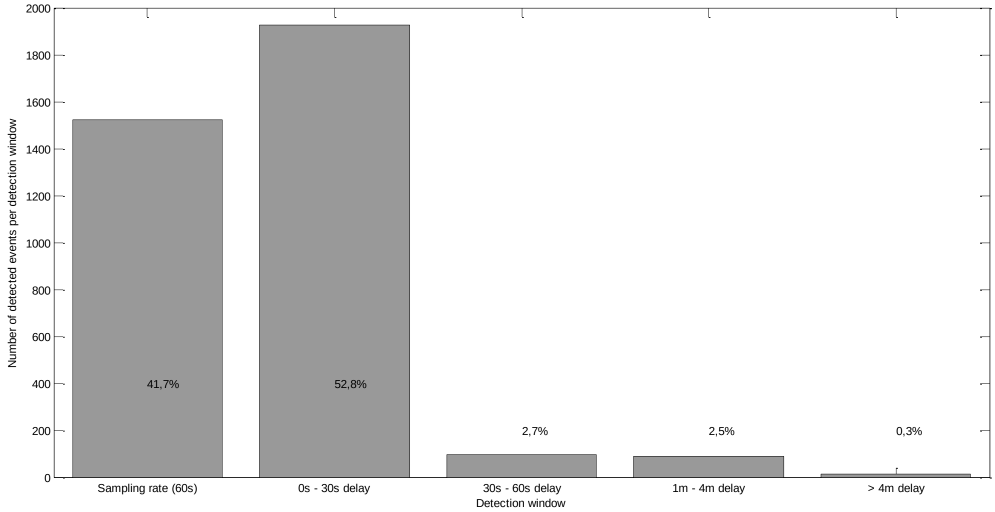


Figure 18: Overview Wi-Fi events logged per detection window

4.3.4 Usability analysis of discussion.

This section shows the results of the face-to-face discussions conducted after the installation process of the AWARE app. The general perception of the installation process was that the instructional video was experienced to be uncomplicated and straightforward. In addition, the installation process was easy. After a feedback session with each participant the next topics were parts that could be improved for future studies:

- The instructional video was too brief in certain parts, especially the section on scanning the QRcode to join the study was too fast;
- The instructional video provided another "Join study" page in comparison to some of the phones of the participants. In the instructional video, the page provides an "OK" and a "quit study" button, and the participants' page provided a "sign up" button;
- A couple of times the app didn't want to start up and crashed. This was because the active button in the AWARE app was activated. This was not instructed in the video, but on the smartphone, in the video, this was already activated. Restarting or reinstalling the app had no effect on fixing it. Even when all the files were removed using a file manager application, it did not work. It only worked when the phones were reset using the "Factory data reset " button, see Figure 19 It shows that the active button on the top should not be active.

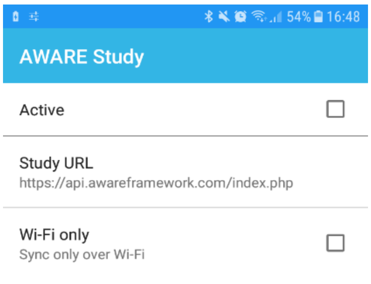


Figure 19: AWARE study - screen with active button

- The QRcode would not scan each time the scanner on the app was activated. This was solved by rotating the smartphone horizontally and scanning the QRcode again.

4.4 Discussion

In what follows the results of the separately evaluated parts have been discussed.

4.4.1 Evaluation results battery drainage and Wi-Fi sensor

The results of the battery drainage analysis showed that the AWARE app drains the battery the battery between 40% to 60% over a 24-hour period. These are the results of not using the phone or having other apps (with the addition of Paco) installed and used. In this 24-hour period, the smartphone was not charged. Therefore, when participating in mobile data sensing studies the battery should be charged at least once a day.
The Wi-Fi sensor with which location tracking causes some minor jitter in the sensor data. 41,7% of all logged detection were collected within the sampling rate and 52,8% was detection the delay detection window of 0 to 30 seconds. These are good results. For future research, it was decided to increase the detection period to every 30 seconds. This is to ensure any change in whereabouts will be captured.

4.4.2 Usability evaluation of discussion

The discussion conducted with each participant resulted in interesting findings. First, the installation video was perceived as straightforward and uncomplicated. The installation process was experienced as easy. These are very positive results. Moreover, some suggestions were made to improve the installation process and installation video.
A few suggestions have been made to improve the installation video to make the experience to watch it more pleasant and even more clear. First, to make the explanation more elaborate, specifically the part of scanning the QRcode to join the study. Second, to shoot a new video with a clean camera from the university. Third, students would like to see a more elaborate instruction vid. Specifically, the part of scanning the QRcode to join a study. A suggestion was done to re-do the installation video with a clean camera from the University. In this new video should show a “Join study” page with a sign-up button instead of an “OK” & “quit study” button. This difference probably has to do with the version of the AWARE app. Fourth, the ‘activate’ button will be disabled as it causes the application to crash for a lot of participants. Lastly, when the QRcode will not scan to join the study, the smartphone can be turned horizontally. This should be mentioned in the new video.

4.5 Conclusion

The last step of this chapter, the sub-objectives as defined at the beginning of the chapter are examined. In what follows the sub-objectives will be discussed:

O21: To assess the usability of the system.

In general, the opinions with respect to the installation process of AWARE was positive, the video was easy to follow and straightforward. In addition, the installation process was easy. Of course, with every experiment improvements could be addressed by participants. The next suggestions were made to improve the instructional video:

- More elaborate explanation. Especially of the scanning of the QRcode;
- For the instructional video, a smartphone reset to factory setting must be used;
- The 'active' button in the AWARE app must be deactivated in the video;
- The instructional video will show a "Join study" page with a sign-up button instead of an "OK" & "quit study" button;
- The new video should mention that when difficulty is perceived scanning the QRcode, that the phone should be turned horizontally;

O22: To investigate the battery drainage caused by AWARE.

From the results, it became clear that AWARE does drain the battery. The drainage is between 40% and 60% over a period of 24 hours without charging the phone. Keeping in mind that in this experiment just the AWARE drainage was investigated. It should be taken into account that for future research on regular smartphones, the phones should be charged at least once a day.

O23: To investigate if there is jitter in the Wi-Fi sensor data.

Minor jitter was detected which means that no large data gaps were detected. This means that a majority of the collected data was detected within the sampling period or within the first 0 to 30 second of delay. These are good results. For future research, it was decided to increase the detection period to every 30 seconds. This is to ensure any change in whereabouts will be captured.

Reflecting on the main objective, it is concluded that these technical and usability evaluations could be useful for future mobile data sensing studies. This is the last experiment to fulfill the first objective. In the following chapter, an attempt will be made to investigate the willingness of subjects to participate in mobile data sensing studies.

Experiment 3: USERS WILLINGNESS TO PARTICIPATE IN MOBILE DATA SENSING STUDIES

The chapter presents the third experiment which aims to reach the second objective. This is to investigate the willingness of subjects to participate in mobile data sensing campaigns. In this experiment, a modified version of the *mobile stress study* was introduced. To reach this objective, two sub-objectives were investigated:

O31: To investigate which tools can be used to investigate the participatory willingness
This is important to get the best results from a behavioral science perspective instead of an engineering point of view.

O32: To investigate the factors that determine the participatory willingness of subjects for mobile data sensing studies.
It is important to establish these factors as little research is done on this topic. In addition, ascertain these factors aid in the proper recruitment of subjects in mobile data sensing experiments.

From the interview with the domain expert, it became clear that the approaches to reach this objective were semi-structured interviews and online questionnaires.

5.1 Subjects

Subjects (i.e. male/female, ratio unknown) of the University of Twente and other universities, 18 years or older were recruited for this experiment. The subjects were given the e-mail address of the principal investigator and asked to send the principal investigator an e-mail when they wanted to participate in the interviews. They were also given the choice to participate in an online survey.

5.2 Method

This section discussed the modified version of the mobile stress study, privacy and safety, execution of the experiment, and the data collected.

5.2.1 Modified version of *mobile stress study*

The modified *mobile stress study* that was introduced in experiment 3 to the subjects will be explained in this section. This modified version is the version that would be used to reach the stress analysis project mentioned in the introduction. That is to study stress and explore the possibility to derive an empirical model for stress using data collection from smartphone usage data.
The previously *mobile stress study* was supplemented with the Wi-Fi sensor and a Wi-Fi sensor. In addition, the length of the experiment changed from eight days to 50 days. This included 7 weeks of collecting mobile phone usage data through the AWARE app and 50 days of ESM and survey questions through the Paco app. In this modified version both apps would run parallel as well.

5.2.2 Privacy and safety

This experiment involved sensitive subjective data, therefore it was necessary to give an informed consent before being allowed to participate in this experiment. The data collected during the interviews were recorded with a voice recorder. The data was transcribed and every participant

was assigned a number (Subject 1, Subject 2) to ensure anonymity in the study. The data corresponding to each number was saved on a secure part of the UT server that could only be accessed with a password.
For the anonymous survey, the data was stored on the UT server, protected by a password and login. To assure anonymity in this experiment the survey was conducted anonymously. In addition, all audio files were deleted after all the audio files were transcribed.

5.2.3 Data collection

The data for this participation experiment was gathered by conducting an online survey and ten face-to-face interviews that were audio recorded. For the content of the survey, please see Appendix 5. Here it can be seen that the question seven and eight are marked, this represents that these questions only had to be answered when the answer to question seven was yes. If the answer was no, then two questions would be skipped. In addition, the content of the interviews can be seen in Appendix 6.

5.2.4 Execution of experiment

For the experiment, a recruitment pitch was held. During this pitch, the students were asked to participate in an interview (which would be recorded) or to participate in an online survey. The students were provided a URL-link for the survey https://goo.gl/swnPa5 and the e-mail address of the principal investigator. In addition, on the pitch, it was mentioned a lottery would be held under the subjects of the interview as well as for the survey. With the survey, it was possible to win five euro's and with the interviews, it was possible to win ten euro. After the pitch, students were reminded twice, that they could sign up for the experiment.
After recruitment and the subject having signed a consent form, the interviews were conducted. The interviews were recorded with an audio recorder, the interviews lasted between 20 and 40 minutes depending on how much a subject had to tell. The topics of the interview included the perception towards the different mobile sensor data, the length of the experiment, the design of the study, motivation to participate, privacy, and security. These topics were also addressed in the online survey. The survey was conducted online and the results were collected ten days after the last participant filled in the survey. Afterwards, the lotteries were conducted via www.naamloten.nl and the winners were provided with their rewards.

5.3 Results interview

The questions of the survey and the interviews were related to each other. In the analysis, the content of the survey and interviews were analyzed separately. Subsequently, it was determined whether there was any correlation between the answers of the interviews and results of the survey.
In what follows, it will be discussed how subjects' perceptions deviated for the various types of parameters, their attitude towards different aspects of the *mobile stress study*.

5.3.1 Subjects

A short recruitment pitch was held in bachelor courses, the recruitment presentation was held for the modified *mobile stress study*. In addition, participants from different Dutch universities were asked. These universities included:

- Stenden University of applied sciences;
- AHK (Amsterdam);
- Wageningen University;
- Utrecht University;
- Radboud University.

Eventually, ten subjects were recruited for the interviews. Nine participants signed the consent form designed for this interview. However, one participant did not want to be recorded. Therefore, an alternate consent form was drafted which stated no audio recording would be made of the interview. After this, all the interviews were conducted with different students. The demographics of the participants are described in Table 10.

Table 10: Demographic characteristics

Demographic Characteristics		N	%
Gender	Female	9	90
	Male	1	10
Nationality	Dutch	8	80
	Chinese	1	10
	Pakistani	1	10
Age	20 - 24	6	60
	25 - 29	4	40
Education	University	8	80
	University of Applied Sciences	2	20
Bachelor/Master	Bachelor	3	30
	Master	7	70
Place of education	Enschede	4	40
	Amsterdam	1	10
	Wageningen	1	10
	Utrecht	2	20
	Emmen	1	10
	Nijmegen	1	10

5.3.2 General opinion

In general, the participants considered the *mobile stress study* to be interesting and complex, but it also issued concerns with respect to privacy and that much was asked of the students. Subject 3 emphasized, privacy issues and putting too much effort in without getting anything in return as the main reasons not to participate. This was her conclusion from a discussion she had with fellow students after the recruitment presentation.
Surprisingly, an overwhelming majority would participate, if they would know the investigator performing the experiment. If this would not have been the case, the experiment was thought of as too private and sensitive.

5.3.3 Responses to the different types of parameters

Table 11 shows that the opinions concerning the various parameters differ amongst the participants. None of the subjects had any regards toward device usage, as it was considered not very sensitive or in other words *"Because no real personal information can be traced back"* {Subject S5}. In addition, the 90% of the participants expressed to have no concerns with the detection of battery life, physical activity, or application use. Battery life and physical activity were believed to be insensitive information. Application use was received a bit differently, even though 90% of the participant expressed no real concern. App name detection was considered more sensitive in comparison to battery life or physical activity. However, as long as just the app names and no content would be detected, it was not considered to be a problem. The opinions for the participants varied depending on the kind of applications on their phones. One commented *"It becomes a bit more private, however, I don't have very exciting apps on my phone"* {S10}, and another one clarified *"It is not as if I use weird apps on my phone, so that saves me."* {S6}.

Reactions to the detection of calls and messages were varied. 60% (6 of 10) of subjects expressed no real concerns for the detection of this parameter. Taking into consideration that no content would be collected and that the contacts would stay anonymous, one noted *"If it is all anonymous then I don't mind"* {S6}. Contrariwise, some concerns were uttered such as *"To be honest, I find it a bit scary"* {S7} or *"I wouldn't consider calls to be a problem, but I would find messages annoying"* {S8}. Some participants considered the detection of calls and messages to be a burden. When asked why they did not opt any concern for the collection of app names such as *WhatsApp*, one marked *"Not that you have anything to hide, but this comes very close to your daily life"* {S7}.
Bluetooth detection was considered not to be a concern by 60% of the participants. Concerns were mostly uttered with regard towards the people whose devices were detected using the Bluetooth on the participant's devices. Other reasons were that the parameter was thought to be weird or unnecessary. *"I can give you my information with my consent, but I am not sure about my surrounding and about the people you knew I hang out with"* {S1}.

Tracking location was regarded as an exception as can be seen from Table 11. A majority of 60% of the participants revealed their agitation towards location detection. This information was considered sensitive and private, and feelings of unease were expressed. Different notes were made from *"That I find somewhat sensitive"* {S6} to *"Not everyone has to know where I am every moment of the day"* {S2} and *"No, would be a bit uncomfortable"* {S10}. The subjects who didn't have a problem with the detection of location argued that they were being watched anyway. *"I am being watched anyway"* {S1} and *"Anonymized for science and a dataset that is anonymized etcetera… And that actually… that are but a relatively short period of time in my life, I don't see a problem in that"* {S9}. When asking the participants if they would like to choose their own parameters for such an experiment, the answers was a unanimously: *"yes"*. It would give more autonomy about your own data, one said *"I think it is a pretty interesting offer"* {S1} and a second one voiced *"Yes than you have a little bit of control over it…."* {S8}.

Table 11: Response overview of mobile stress study parameters (N=10) - Results from interview

Measurement	Sensors & plugins	No problem	Problem	Not sure
Physical activity	Activity recognition	9	1	
App use	Application sensor	9	1	
Device use	Device usage	10		
Calls & messages	Communication sensor	6	3	1
Surrounding devices	Bluetooth sensor	6	4	
Battery life	Battery sensor	9	1	
Location	Wi-Fi sensor	4	6	

5.3.4 Responses towards ESM and length of experiment

Unlike the distribution of the parameters mentioned above, all participants agreed that experience sampling that happened six times a day for seven weeks was too much. Most participants argued that the frequency of questions should decrease to two or three times a day (60%). Surprisingly, the majority was prepared to answer more than two questions at a time. One commented *"Yes, maybe two times a day… then there might be more questions."* {S6}, another comment was *"So, then it does not matter if it is three or eight"* {S5}.
The opinions on the length of the experiment were divided. Everyone agreed that something needed to be changed, which means either the length of the experiment, the frequency of

questions asked during the day or both. 70% of the subjects understood that the length of the experiment was necessary to reach hypothetical *mobile stress study* as discussed in the methods. However, a majority either thought this was still too long or that seven weeks was more acceptable if the ESM frequency would be decreased. A majority (70%) of the students thought a reduction of the experiment length would be necessary and doable, this time ranges from three to five weeks*. "I don't think that I can last for seven weeks"* {S5}, *"I think it would be easy to eventually forget to fill it in"* {S8}, and *"… the frequency is more of a concern"* {S4} were used as arguments for the length of the experiment.

Security & privacy

The participants were asked about their thoughts on anonymity with respect to the encrypted IDs used for the participant's smartphones and the contacts in their smartphones. Everyone agreed that this was an important feature. One subject articulated "*I think it is important. Nowadays, I think less is not acceptable as it is so easy to do"* {S10}.
A majority of 90% of the interviewees did not express concerns about sending the data, collected on the smartphones, encrypted over a Wi-Fi connection to the UT server.
Something considered to be a concern was the messages send by Android as can be seen in Appendix 1, that appear during the installation process of AWARE. Most students either would not agree or would have hesitations if they would see the messages. However, 40% still would agree with the terms of these messages. *"It makes me a bit hesitant, I think… "*{S7} voiced one, another one said *"you get this so often and actually I always just press accept"* {S10}.

Motivation and changes

It was discussed with the participants how they could be motivated and what they would like to see changed in the design of the experiment to reconsider joining the *mobile stress study*. Overall, students got the impression that a lot was asked of them without getting anything in return. When asked what they would like in return, the answers varied from either wanting to receive a reward in terms of money (50%) or wanting to see the data from their smartphone. 90% of the participants wanted to see their data either during or at the end of the experiment. 66,7% of these participants thought to see the data would motivate them more to participate in the experiment. One participant articulated *"At the end of the experiment, it would be interesting to see …. What I actually do with my phone"* {S8} and another *"So, I would like a good visual representation, like a general representation that a layman could understand"* {S10}.

Other motivational factors included: messaging prospective participants over and over again to remind and familiarize them with the experiment and design a coaching system with the stress model.
In order to reconsider joining the *mobile stress study*, the student mentioned some concepts that they would like to see changed. Most participants agreed on the following changes: a visual representation of the data, being able to choose the parameters measured, decreased frequency in which questions are asked, and a smaller period in which the experiment could be conducted.
Lastly, participants thought the presentation was a good way to reach a lot of people at once. A few suggestions were made to improve this method of recruitment. First of all, the amount of information seemed a bit daunting all at once. It was suggested to make a short and motivating pitch and just say that everything would be anonymous. This would keep the attention of the students and would not scare them off as easily. It was also suggested to perform the recruitment pitch at the end of a lecture. This could allow students to discuss the experiment immediately after.

5.4 Results online survey

In what follows, the results of the survey will be discussed. For the results, it was decided not to include statistical analysis as the sample size was too small to conduct any significant analysis. Rather, it was analyzed what the responses were towards the statements of the survey.

5.4.1 Subjects

A group of 13 students participated in the *Survey – mobile stress study*. After multiple reminders and 10 days of no response, it was decided to close the survey and analyze the results.

5.4.2 Recruitment presentation

The statements with respect to the recruitment presentation are described in Figure 20. The figure shows that over 60% thought the information given during the presentation was very informative. However, most students had either a negative or neutral opinion regarding the presentation. Moreover, a majority of 70% thought they would not have been motivated more to join the study when encouraged more. In addition, no one of the participants had visited the webpage. Therefore, questions seven and eight on the topic of the webpage and instructional video had not been answered.

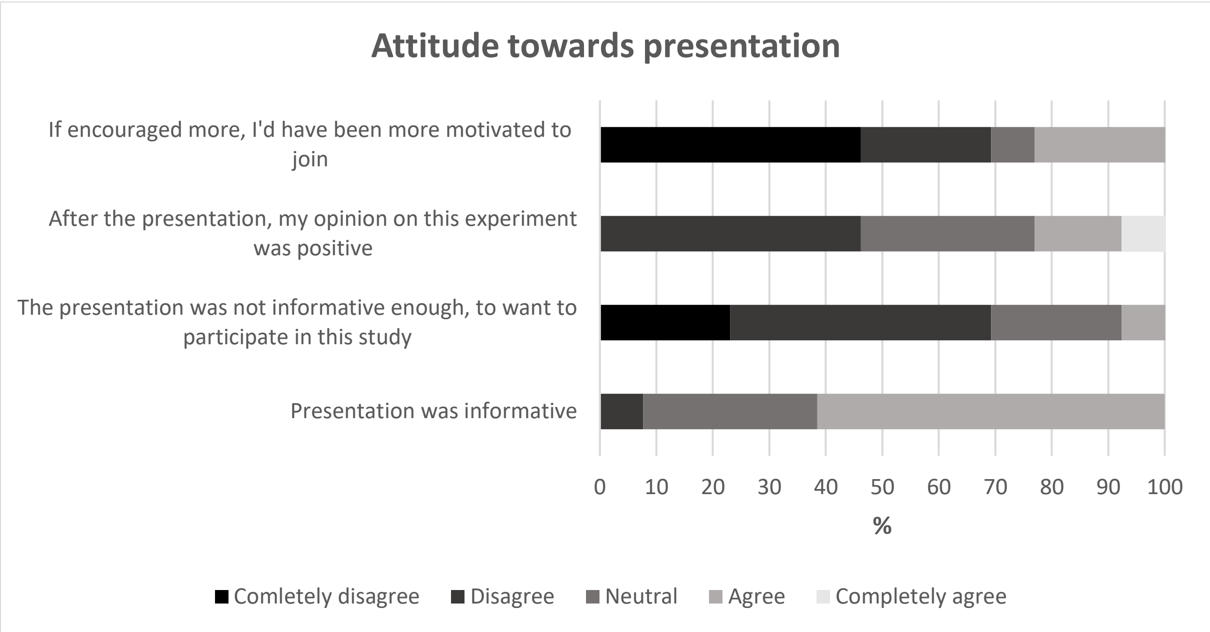


Figure 20: Responses to statements with respect to presentation

5.4.3 Data parameters

Figure 21 shows the opinion measured using a Likert-scale. To start with, participants had different opinions when it came to measuring smartphone data 24/7 or only during certain parts of the day. Over 50% thought that measuring 24/7 would be better, and almost 30% was neutral of opinion. Moreover, more students agreed than disagreed that logging smartphone data all day was too long. However, 30% was neutral in opinion about this.

A majority (~70%) finds location detection unacceptable and just a hand full of approximately 15% thought it was acceptable to track the location of a subject. When asked whether they thought it would be different if the location was monitored using a Wi-Fi sensor in comparison to GPS, participants had deviating opinions.

Over 60% of the subjects thought having fixed times to be asked questions during the day is better in comparison to randomized times. On the topic of questions, most subjects found it to be unacceptable that questions were asked six times a day. Contrariwise, a majority found two questions at a time to be acceptable. It was also agreed to be a good thing to send questions that were based on scientific research.

More than half of the participants believed that monitoring calls and messages were not acceptable. This opinion was also shared on the matter of detecting the contacts involved in text messages and calls.
Logging social media apps were found to be acceptable to over 50% of participants, and about 30% did not have an opinion on this matter. Communication apps, such as WhatsApp, were accepted to be detected by approximately 40% of participants. Another ~40% were neutral towards communication apps to be detected. Bluetooth detection was disliked by over 50% of the participants. In addition, over 50% of the participants did not believe that people surrounded by them would be protected if the Bluetooth sensor would be activated. Lastly, almost 70% of the subjects found the data to be too invasive.

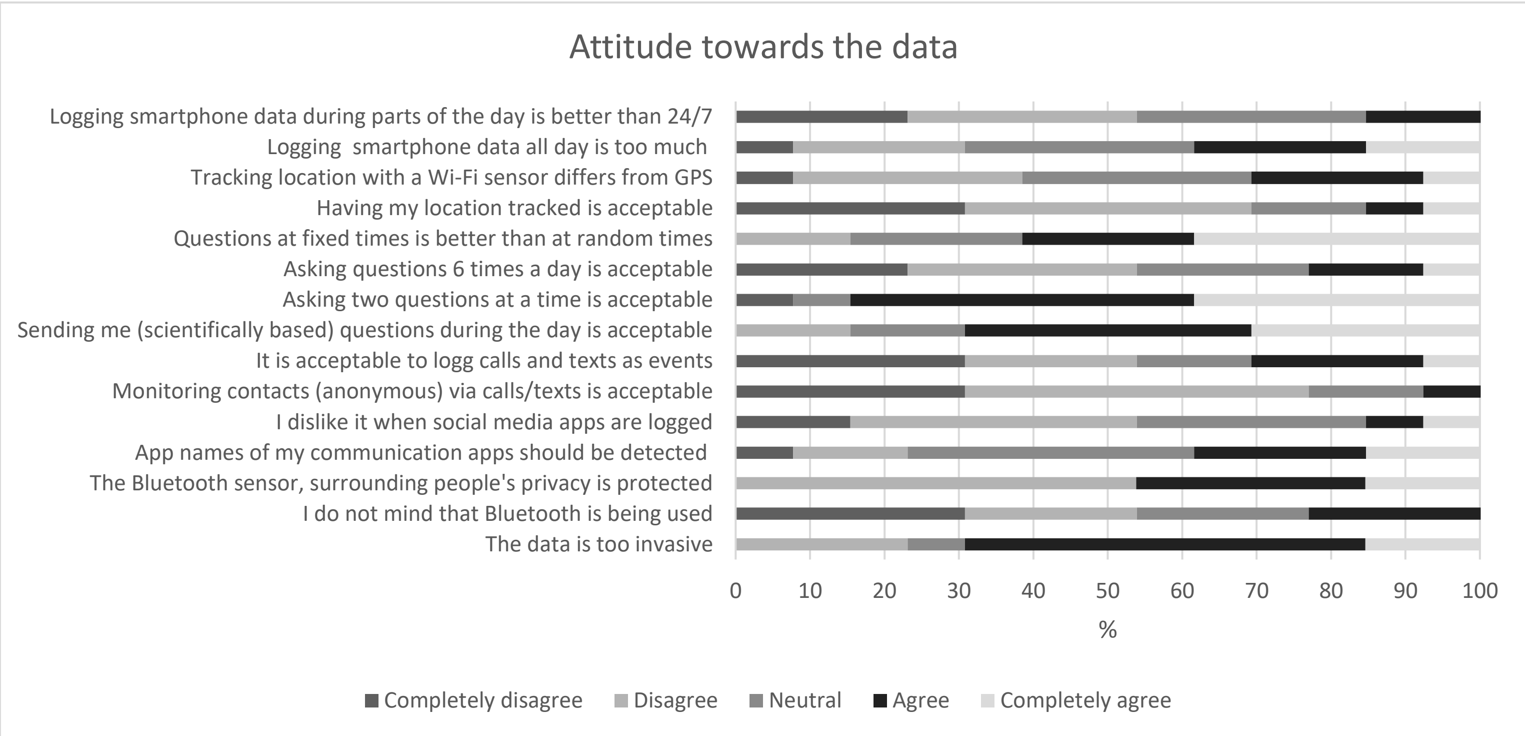


Figure 21: Opinions with respect to parameters of the mobile stress study

As stated above, the subjects found the data measured to be too invasive. Figure 22 shows a ranking which parameters of data collection have been found to be most invasive. The figure shows that detection of battery life and smartphone use have been experienced as least invasive. These are followed by Bluetooth and physical activity. Detecting calls and messages, detecting app use, and questions asked during the day were found to be more invasive. Lastly, over 30% found location detection to be most invasive. Note, 5 participants, filled this item of the survey in properly. Therefore, only these results were used.

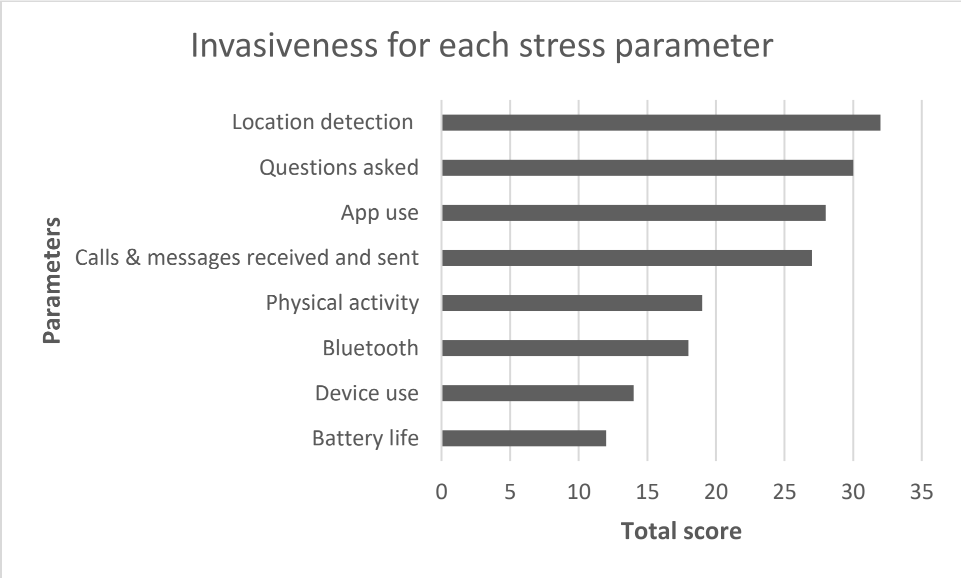


Figure 22: Ranking of invasiveness parameters (N=5)

5.4.4 Security & privacy
Figure 23 shows statements with regard to security and privacy, how they were assessed by the subjects. Approximately 70% of subjects experienced it as safe that collected data was encrypted and send over a Wi-Fi connection the UT server. In addition, it was also found acceptable by over 40% of participants to store this data on the UT server. Another 40% was neutral with respect to this item. Lastly, over 50% believed that an encrypted ID would guarantee the anonymity of a smartphone user.

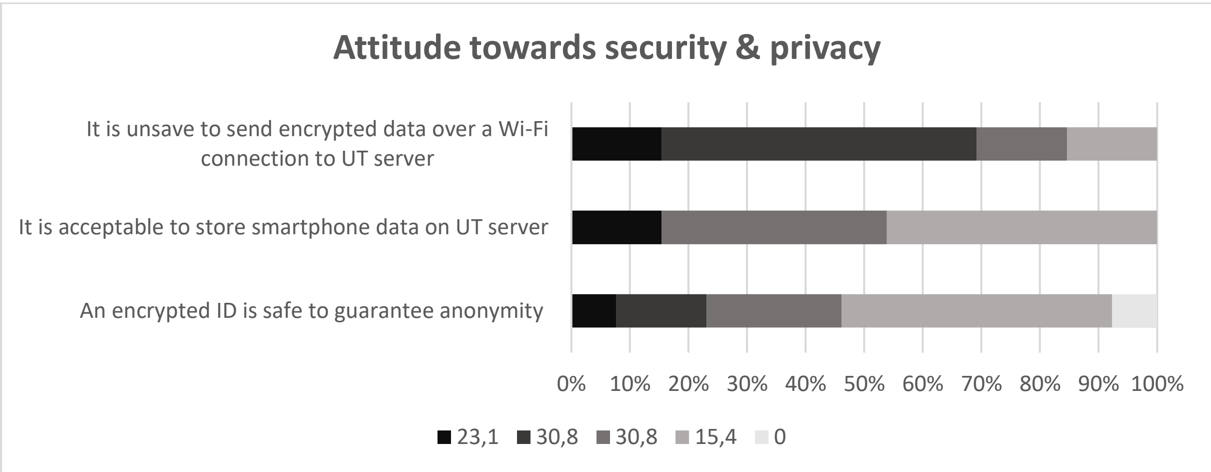


Figure 23: Attitude towards security & privacy statements

5.4.5 Motivation and changes
Figure 24 shows the reasons for students not wanting to participate in the *mobile stress study*. It is clear, that the length of the experiment is the main factor for students not to want to participate. Over 20% thought the experiment was too invasive and another 15% believed the experiment would take too much effort to join.

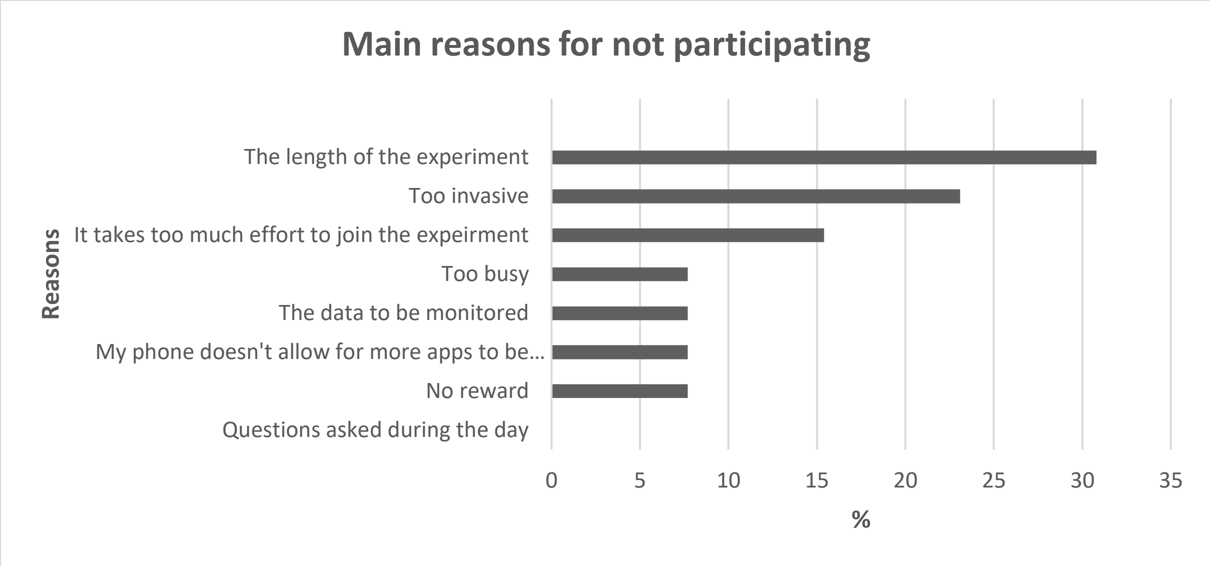


Figure 24: Main reasons for not participating

The survey contained a question that asked what they would like to see changed in the design of the modified *mobile stress study*. The participant shared various reasons, which they wanted to see changed. The next list includes the changes that they would like to see:

- The experiment should be shorter;
- There should be less than six interruptions throughout the day;
- Gaining insight into their own stress levels;
- Making it easier to participate in the experiment;
- Getting something in return for joining the experiment;
- The exclusion of some parameters, being able to choose what can be monitored and what not.

Especially, the first two points were linked to each other. Some participants argued that they would like to see either one changed in order to accept the set experiment length or the set ESM frequency. Figure 25 shows that if the changes mentioned by the participants would be made that over 50% would reconsider joining the experiment. A little over 20% would not want to join the experiment at all, even if changes would have been made. Almost 40% of participants would be motivated by a reward to join the experiment. However, over 40% would not be motivated by a reward.

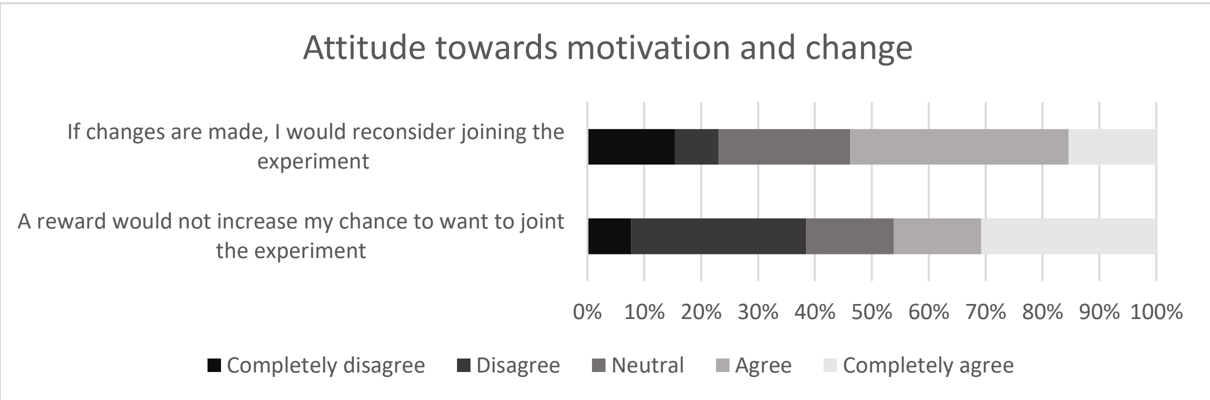


Figure 25: Subjects' opinion with respect towards changes and rewards

5.5 Discussion

In what follows the results of the survey and the interviews will be compared to see if any correlations exist. From this the main reasons for not participating, future changes, and other items will be deduced.
In general, the information provided in the recruitment presentation was experienced as informative. However, the information was also very detailed which caused concerns. It made the experiment seem daunting and very invasive, without getting anything in return. In addition, another concern was that prospective participants don't like to share their sensitive information with a stranger. Especially, if no reward is offered.

5.5.1 Stress parameters

The opinions on the different parameters were divided. Some expressed their concern towards the invasiveness to some parameters, while others did not have a problem with the parameters mentioned in the experiment.

The interviewees expressed the fewest concerns towards monitoring phone use. The participants of the survey put smartphone use detection in second place of least invasive. Battery life was also not mentioned as a big concern for both parties, in the survey, it was voted least invasive. The two parameters were not found to be very invasive, as the participants believed not a lot of personal information could be deducted from it. In addition, physical activity was to be found a good parameter to measure with respect to stress. The interviewees thought this parameter was less invasive in comparison to the survey subjects.

Surprisingly, detection of app names was not a big expression of concern for the interviewees. Although, it was experienced as more private in comparison to the previously mentioned parameters. Conversely, the survey participants expressed some real concern towards the detection of app names.

Bluetooth, as well as detection of calls and messages, caused some concern for the participants. Both parameters were thought of by the participants of the survey to be not very acceptable, over 50% were of this opinion. While the interviewees expressed less concern. 60% of the interviewees expressed that it would be acceptable to monitor these elements. Provided that, the anonymity of contacts was guaranteed and no content would be measured.

An overwhelming majority of both groups found location detection unacceptable. It was experienced as intrusive and ‘big brother’ like to monitor someone all day long. This information considered very private.

Lastly, the frequency of ESM during the day needed to be decreased. Six times a day was found to be very invasive and intrusive. However, it would be more acceptable if the frequency would be decreased to two or three times a day. The amount of two questions at a time was expressed to be acceptable. A few interviewees mentioned that if they were answering questions already, it would not be a problem to answer a few more. However, this amount of questions could not go higher than eight. Most of the participants preferred five questions in this case.

5.5.2 Security and privacy

Everyone agreed that measures to ensure anonymity and privacy of the collected data were important. Of the interviews, no one expressed concerns using a special ID for anonymity purposes. Over 50% believed that the ID would guarantee the anonymity of a smartphone user.

In addition, most participants (70% survey and 90% interviewees) believed that the encrypted data would be sent safely over a Wi-Fi connection to the UT server. Storage on the UT server gave a bit of a discussion, some participants don't believe it is safe. While others didn't like it that their data could be accessed by other University scientists in the future.
The messages (Appendix 1) sent by Android during the installation process could cause a bit of a concern in the future. However, little can be done about this as it is a fixed part of Android's installation process.

5.5.3 Main factors and improvements

Throughout the experiment, a variety of reasons were offered for students not wanting to participate in the *mobile stress study*. Striking is that most arguments are correlated with each other. Almost every participant thought the length of the experiment was too long. In addition, the frequency of ESMs during the day was too high. Decreasing either one of these factors would be helpful to get more participants to join the study. Most participants expressed that they would like to see a reduction of ESM frequency and a decreased experiment length. Additionally, participants indicated that they would not mind answering more questions at a time if the frequency of ESM were lower

Too much effort and too little in return was something heard quite often during the interviews. Participants would like to see this changed in the form of a reward. This reward differed between the participants, some wanted a reward in the form of money and others would just like a visualization of the data collected from their phone.
The amount of data and the sensitivity of the data was a big problem for most participants of this experiment. However, students would be more willing to participate when able to choose whether or not parameters are activated. Lastly, something voiced less frequent was the amount of effort it took to join the experiment.

5.5.4 Limitations

The number of participants in this experiment was still low to be able to draw any statistically significant conclusion.

5.6 Conclusion

The final step of this chapter, the sub-objectives as defined at the beginning of the chapter are examined. As previously mentioned to fulfill the objectives of this thesis, the thesis is divided into three experiments each experiment has its own sub-objectives. In what follows the sub-objectives of experiment 3 will be discussed.

O31: To investigate which tools can be used to investigate the participatory willingness

According to literature research and an interview with the domain expert, the best tools to investigate participatory willingness is with interviews.

O32: To investigate the factors that determine the participatory willingness of subjects for mobile data sensing studies.

The arguments mentioned in the results and discussion lead to a set of factors that would improve the willingness of students to participate. Implementing these factors could lead to more participants for the *mobile stress study* and future experiments. The following guidelines could be implemented to improve the participatory willingness:

- Experiments should last from three to five weeks;
- ESM during the day should ideally be two to three times a day at fixed times;
- The number of questions for each ESM should vary between two to five times;
- Participants must be rewarded for mobile data sensing studies. For instance, by providing students with a visualization of their own data. Depending on the amount of data to be collected and the length of the experiment a monetary reward should be considered;
- Information retrieved from phones is very sensitive and private. Therefore, the option to let prospective subjects choose data from a list must be considered;
- To motivate prospective participants, an attractive and motivating pitch must be used as a recruitment tool;
- Only one app should be used in mobile data sensing studies to simplify the process of joining an experiment.

Reflecting on the main objective, it is concluded that these factors that students would like to see to be willing to participate, could be useful for future mobile data sensing studies. The target audience in this experiment were students across different levels of education (i.e. bachelor or master). In addition, students were interviewed from different Dutch universities. Therefore, it can be argued that these factors that determine the willingness of subjects to participate in mobile data sensing studies can be generalized for all Dutch universities. However, further research might be needed as no statistical analysis could be performed in this experiment.

This was the experiment to fulfill the second objective. The following chapters include discussion across all three experiments, a conclusion, and recommendations.

Discussion

This chapter has the purpose to discuss the major findings of the thesis, to explain the meaning of these results and their importance, to relate the findings to the discoveries of similar studies.

6.1 Summary and experiment findings

This thesis started to accomplish two objectives. The first objective was to investigate the technical and usability aspects relevant for an effective mobile data sensing campaign. The second objective was to investigate the willingness of subjects to participate in mobile data sensing campaigns. To fulfill the aim of the first thesis objective a *mobile stress study* was introduced, to help investigate the possible jitter in sensor data (through AWARE) and to investigate the response of students with respect to the questions asked (through Paco). In addition, the usability of both apps and the experience of the experiment by the participants was investigated. First, the parameters relevant to stress that could be linked to the available sensors of AWARE involved are physical activity, and social support/interaction. Additionally, subjective data was deemed important as not all data relevant to stress can be measured by smartphone use (through questions via Paco app). The sensors of AWARE used in the experiment were:

- Activity recognition plugin: to detect activity events
- Application sensor: to detect which apps are being used
- Communication sensor: to detect calls and text events
- Device usage plugin: to detect how often a smartphone is being used
- Bluetooth sensor: to detect surrounding devices as a measure of isolation
- Wi-Fi sensor: to detect location (introduced in the second experiment)

The possible jitter in the sensor data of the activity recognition plugin, application sensor, Wi-Fi sensor, and the Bluetooth sensor was small. This means, for all these sensors that over 75% of data was collected within the set sampling rate or within the first 0 to 30-second delay detection window. It suggests the indication that due to a minor jitter in the data, the quality of the data is good. The device usage and communication sensors generate data asynchronously thus no sampling rate applies here. This suggests another approach to see if the sensors collect the data accurately. The communication sensor showed that every time a call or a text event took place, that the event was logged. The device usage data showed two different ways the timestamps were collected. One, that during the time a smartphone was locked or unlocked the seconds were collected and in the analysis added together. This was done to see how long the data collection had taken place. Second, all detections of the device locked or unlocked were accompanied by a time stamp. The first and last time stamp were taken and the difference was calculated, to see how long the data collection had taken place. These two calculated time stamps were compared with each other and show for each participant a small difference of a couple of minutes. This suggests that only one of the aspects should be sued for future research.

The investigation with respect to the Paco app showed that the average response time to questions was 5.30 minutes. In addition, 88% of answered questions were responded before a reminder was sent to the smartphone after 15 minutes. Also, most questions were answered in the afternoon. Lastly, 74% of all questions send to the participants were answered which looks like a promising result.

The usability was investigated with a survey at the end of the first experiment and with a face-to-face discussion after the installation process of AWARE. The survey was answered by one subject, meaning that the results cannot be generalized. However, some interesting findings were extracted from the results of the survey. These findings include the SUS score for the *mobile stress* *study* to be 70, in words this means acceptable. The overall friendliness score is OK, which suggests that some improvements might be fruitful for the experience of the *mobile stress study*. To improve the usability the survey questions should be available for longer on the smartphones, so no questions could be missed and no vibrate setting has to be used. It should be more emphasized during recruitment that the questions asked during the experiment are based on scientific research. The usability was assessed with respect to the AWARE app installation, which was experienced as pleasant and easy. As for any research, some improvement was suggested for a new installation video to make it clearer, easier, and to prevent negative implications for future research.

To fulfill the second objective of this thesis, the participatory willingness in mobile data sensing studies was investigated with surveys and interviews. The questions were developed with the introduction of a modified version of the *mobile stress study*. This experiment was conducted using students from various Dutch universities. The results suggested some factors that determine the users' willingness to participate in mobile data sensing experiments. This entails factors such as: acceptable experiment duration, favourable frequency of questions asked during the day, visualization of data collection for each participant, downsizing the project to one application, using motivational pitches to recruit future participants, offering a reward to participating students, and to let participants choose the data to be collected from their mobile devices.

6.2 Meaning and importance of findings

These findings are important because as mentioned in the introduction, this work contributes to a larger project that intends to analyze stress. The importance of the technical and usability findings is to be able to design/modify the *mobile stress study* such that no technical implications will occur during future research. In addition, to make the experience of prospective students as pleasant as possible, by being able to optimize the usability of the mobile data sensing platform. These results have the same implications for other future mobile data sensing studies involving AWARE and or Paco.

Furthermore, assessing the willingness of students to participate in mobile data sensing studies was never investigated before. The literature shows little research in participatory willingness studies, the research that has been done was not specifically focused on students. In addition, the research suggested that the participants were already less willing to participate when installation of an app was involved [58, 59]. The willingness assessment included topics such as the different data, experiment length, ESM frequency if students understand the implications of participating in a mobile data sensing experiment. This gives specific insides that help for future research.

The technical findings, usability discoveries, and participatory willingness findings have meaning to improve the *mobile stress study*. In addition, future data sensing campaigns will benefit greatly from all these findings. The reason for this is that these findings can be used to design a proper mobile data sensing study from the start.

6.3 Relating findings to similar studies

The student life experiment by Wang et al used a mobile data sensing app to investigate the workload of students on stress, activity, sleep, and academic performance across a 10-week term [40]. They used activity, conversation, sleep and location as behavioral classifiers to assess stress with mobile data sensing. This is comparable to this master thesis, however there a couple of differences. For example, the StudentLife experiment activity is assessed using an accelerometer and GPS. A positive aspect of this master thesis is that no GPS was used. The recognition plugin was able to detect the difference between, walking, biking, and sitting. In addition, it was chosen to use GPS for location tracking in the StudentLife experiment, in this master thesis that was done using a Wi-Fi sensor.

In contrast to the StudentLife experiment, it was chosen to not use conversation as an aspect to measure stress with mobile data sensing. Rather it was chosen to collect data of social media app, communication app, and calls and text event use. This was used as a measure of social interaction instead of using conversation what was measured using a microphone. Collecting data with a microphone sensor was suggested to be invasive and sensitive by Klasnja et al [64].
Moreover, both studies used Bluetooth. However, Bluetooth in the StudentLife experiment was used to track location together with GPS. In this master thesis, Bluetooth was used to detect whether students were alone or in the presence of other people.

Lastly, the collection of objective data was performed in both studies. In this master thesis, Paco was used to ask objective data with respect to stress, mood, physical activity, social interaction, and sleep. The StudentLife experiment used Paco as well with some extension features and used SurveyMonkey for pre- and post-experiment surveys. This also included questions with respect to stress, mood, exercise, etcetera. As stated before, this has some similar aspects. However, in this master thesis, only one app was used instead of two apps. In addition, the *mobile stress study* (of this work) includes the GARS scale which assesses current stress and can be asked every day. StudentLife uses the Perceived Stress Scale, which assesses stress with questions with respect to the last month [63].
There have been few studies performed concerning the willingness of students to participate in mobile data sensing studies. However, there have been a couple of studies that have researched some aspects of the interview topics. In what follows these studies will be compared with the findings of experiment 3. A study conducted by Gustarini et al suggested that location monitoring even anonymously was still being seen as a compromise on the privacy of participants [65]. This resulted from experiment 3 as well. A majority of 60% of participants had problems with location tracking, the subjects were uncomfortable with the idea of being watched.

The previously mentioned study by Klasnja et al did also suggest that most participants did not have problems with activity monitoring. This resulted from the interviews as well, for which 90% did not mind if the activity would be monitored. The study from Klasnja is different from experiment 3, in a way that it only has investigated what participants thought of physical activity monitoring, GPS, and microphone data sensing [64].

Harari et al conducted an experiment which suggests that depending on the intrusiveness the willingness to participate in behavioral tracking experiments varies [66]. It investigated aspects such as survey questions during the day, and social media use detection. For example, 54% of participants in the study did not mind if questions would be asked twice a day. In this thesis, 60% did not mind receiving questions twice a day, in both experiments over half of the participants agreed with this. In addition, in Harari's study, it was only asked if participants would provide access to smartphone sensors in general, 46% would do this. This master thesis asked students in the interviews and in the survey of every senor and plugin what their thoughts were on the specific sensor. The studies mentioned above are all the studies that show similarities and differences in results with the findings of this thesis.

Conclusion and recommendations

This thesis proposes to investigate the technical and usability aspects of a mobile data sensing study and to investigate the willingness of subjects to participate in mobile data sensing studies. As a final step, the research objectives as outlined in the introduction will be discussed in this last chapter. In addition, recommendations for further research, and study limitations will be examined. This research as any other research has some limitations. However, the limitations help to identify further research.

7.1 Research objectives

In the introduction chapter, the main research objectives were defined as:

- *To investigate the technical and usability aspects relevant for an effective mobile data sensing study.*
- *To investigate the willingness of subjects to participate in mobile data sensing campaigns.*

To be able to answer these research objectives, a decomposition of the main objectives have been applied. This results in various sub-objectives. In what follows the sub-objectives will be reflected.

O11: To determine parameters relevant to stress and to translate these parameters to sensors

Through literary study, the parameters relevant to stress were identified. Various parameters have been studied. Moreover, not every parameter can be used in combination with mobile data sensing. The parameters identified are as follows:

- Physical activity
- Social support/interaction
- Subjective information

Parameters relevant to stress have been used to identify plugins and sensors of AWARE to collect mobile usage data. It was chosen that physical activity was measured using the activity recognition plugin. Social interaction and support have been combined with the following sensors and plugins available on AWARE:

- Communication sensor, to see how often a subject is being called, calls, sends a text, or is being sent a text;
- Device usage plugin, to see how often a subject uses his or her smartphone;
- Bluetooth sensor, which can be used as a detection method for isolation analysis.
- Application sensor, to see what apps the subject uses. This with a specific interest in the social media apps or communication apps such as WhatsApp.

Meanwhile, the subjective data could be obtained using ESM and questionnaires using Paco.

O12: To investigate the jitter in sensor data

The jitter in sensor data was small for the Bluetooth sensor, the activity recognition, and the application sensor was good, as over 75% of data was collected either within the sampling period or in the first 0 to 30-second delay. This suggests that the sampling data of these sensors is consistent. In addition, the jitter in the sensor data of the device usage plugin and the communication sensor could not be tested using the sampling period as a sampling period could not have been set. However, it was checked that this sensor does measure and collect occurring activities. For that, they work well. Table 12, for performance overview of the sensors.

Table 12: Overview performance sensors

Sensor/plugin	Detection	Remarks
Device usage plugin	Cannot be measured in terms of the sampling period. The sensor performs well.	
Communication sensor	Cannot be measured in terms of the sampling period. The sensor performs well.	
Application sensor	Good	30 second sampling period should be turned into a 90 second sampling period
Activity recognition plugin	Good	60 second sampling period should be turned into a 30 second sampling period
Bluetooth sensor	Good	Bluetooth should be enabled manually
Wi-Fi sensor	Good	60 second sampling period should be turned into a 30 second sampling period

O13: To investigate the response of the participants with respect to the questions sent via the Paco application

The average response time to questions that have been sending by Paco was 5.30 minutes, and 88% of the answered questions have been answered before the first reminder at the 15-minute mark. This is a very interesting result as it suggests that when students participate in mobile data sensing studies, that they answer very quickly. The answering rate amongst the subjects was pretty high. 74% of all sent questions were answered.

77% of the questions send between 12:00 and 15:00 were answered and also 77% of the questions sent between 15:00 and 18:00 were answered. From 9:00 to 12:00 and from 19:00 to 22:00 the response rate was 70%.

O14: To investigate the usability of the Paco app, the AWARE app, and how the experiment was experienced by the subjects.

The SUS score from the *post-experiment questionnaire* was 70 which means that the usability of the system is acceptable. The system was experienced as easy to use, and easy to learn. In addition, the user-friendliness according to SUS was given a score of OK. This suggests that although the study is fine some improvements might be proposed for future research. Examples of future improvements are:

- Displaying the notifications from Paco for a larger time span on the smartphone of the participants.
- Although the questions on stress are related to scientific evidence, it has been experienced that the questions should be more relevant to stress.

O21: To assess the usability of the system.

Positive opinions:

- The installation video was easy to follow and straightforward;
- The installation process of AWARE was easy;

Improvement suggestions:

- More elaborate explanation. Especially of the scanning of the QRcode;
- For the instructional video, a smartphone reset to factory setting must be used;
- The 'active' button in the AWARE app must be deactivated in the video;
- The instructional video will show a "Join study" page with a sign-up button instead of an "OK" & "quit study" button;
- The new video should mention that when some difficulty is perceived scanning the QRcode, that the phone should be turned horizontally.

O22: To investigate the battery drainage caused by AWARE

AWARE drains the smartphone battery between 40% and 60% over a period of 24 hours without charging the phone. Keeping in mind that in this experiment the phones used during the experiment were not used in daily life (i.e. a university smartphone was used). This causes a difference in battery drainage in comparison to daily life smartphone usage. Regular phone charging of at least once a day should be done to avoid large gaps in data collection in future experiments.

O23: To investigate if there is jitter in the Wi-Fi sensor sampling data.

The possible jitter of the Wi-Fi sensor data was small. A majority of the collected data was detected within the sampling rate or within the first 0 to 30 second of delay. These are good results. For future research, it was decided to increase the detection period to every 30 seconds. This is to ensure any change in whereabouts will be captured.

O31: To investigate which tools can be used to investigate the participatory willingness

The theoretical background information in chapter 2 and the interview with the domain expert show that the best tool to research participatory willingness is a semi-structured interview.

O32: To investigate the factors that determine the participatory willingness of subjects for mobile data sensing studies.

The following guidelines could be implemented to improve the participatory willingness:

- One app would simplify the process of joining a mobile data sensing study;
- Experience sampling would be ideal at fixed times each day with a frequency of two to three times;
- A reward for participants would be appreciated. This can be done in different ways, examples are visualization of data or being awarded money;
- The duration of a study can last between three and five weeks;
- Make a motivating pitch as a recruiting tool;
- Every time experience sampling is used three to five questions can be asked;
- Choosing from a list of parameters would make it more attractive to join a mobile data sensing study as sensitive and private information is retrieved from phones.

Reflecting on the research objectives, it can be concluded that the technical aspects, usability aspects mentioned above are relevant for effective mobile data sensing campaigns. The aspects discussed above either are already relevant and effective or with some improvements can be made more effective. In addition, the factors with regard to participatory willingness will contribute towards an effective mobile data sensing campaign. This master thesis can be used as a reference work towards for the intended larger project. This is to study stress and explore the possibility to derive an empirical model for stress using data collection from smartphone usage data. In addition, it can be used as a reference work for future mobile data sensing studies as well.

7.2 Limitations

First, the number of participants that participated in experiment 1. Three participants cannot be used to make significant or general claims with respect to the technical and usability aspects of a mobile data sensing platform such as AWARE. Future research might need to recruit students more intensively. This means, for example, recruiting students from the entire university instead of a couple of courses (consisting of approximately 400 students), or maybe reach out to multiple universities. Another suggestion to consider is rewarding students, as previously mentioned in experiment 3. Students like to be reimbursed for their efforts either with money or being able to see the results of their own data.
The same goes for the 13 participants who filled out the online survey. If recruitment would have taken place across multiple universities or more courses, students might have been more willing to participate. This might be a suggestion for future research. Even though, this is a limitation of this thesis. It should be stated that the results from this thesis include some promising and interesting findings.
Lastly, the results from this current research are relevant towards the target audience of students. If for future research a different target audience will be used, it should be stated that more research might be needed to investigate the specific target audience in mind.

7.3 Future research

Further research can improve the results of the current research, and help to get more generalized and significant results for investigating technical aspects and usability aspects relevant towards an effective mobile data sensing study and for investigating the participatory willingness of students.
The studies could be performed once more, but use larger sample sizes by recruiting students from more Dutch universities, both bachelor and master classes. It takes time to perform such experiments, but it can help to get more significant results. With that new research, a project to analyze stress can be performed properly in the future can be performed.

7.4 Recommendations

Over the course of this thesis, a lot of knowledge was gained with respect to mobile data sensing studies and toward the willingness of students to participate in such campaigns. The aspects learned can be implemented in future research to design a proper mobile data sensing study from the start and aid in recruitment. In what follows recommendations for prospective future research have been listed:

- Experiment duration: three to five weeks;
- Questions during the day: two to three times at fixed times;
- Every time it is possible to ask two to five questions;
- Participants should be rewarded;
- Data collected from participants should be visualized for the subject to see;
- A motivating and attractive pitch should be used to attract prospective subjects;
- Using one app in a mobile data sensing study should simplify the process of participating;
- Participants should be able to choose what can data can be collected from their smartphones;
- Questions should be available for longer on smartphones, which means longer than half an hour. This in order to not miss questions and not to use the vibrate setting on the phone
- Let participants charge their phone at least once a day as AWARE can drain the phone. It does, however, depends on how much a smartphone is used by the participant;
- Record an installation video for the AWARE app that is elaborate and explains everything slowly to minimize installation mistakes.

References

[1] Seltzer LF. “Trauma and the Freeze Response: Good, Bad or Both?” Retrieved from *https://www.psychologytoday.com/blog/evolution-the-self/201507/trauma-and-the-freeze-response-good-bad-or-both* (2015)
[2] Fight stress with healthy habits. American Heart Association. Retrieved from http://www.heart.org/HEARTORG/HealthyLiving/StressManagement/FightStressWithHealthyHabits/Fight-Stress-with-Healthy-Habits_UCM_307992_Article.jsp#.Vr5VAtj2bcs (2015)
[3] How stress affects your health. American Psychological Association. Retrieved from http://www.apa.org/helpcenter/stress.aspx (2013)
[4] Juster R, Martin M. “Genetics and Stress: Is there a Link?” *Mammothmagazine* (2011)
[5] http://www.juniordentist.com/wp-content/uploads/2012/09/Pituitary-gland-anatomy.jpg
[6] Rizzo, Donald C. *Fundamentals of anatomy and physiology*. Cengage Learning, 2015.
[7] Porth, Carol M. *Pathophysiology: Concepts of Altered Health States. 5th*. Philadelphia, Pa, USA: Lippincot-Raven Publishers, 1998.
[8] Cohen, Sheldon, T. Kamarck, and R. Mermelstein. "Perceived stress scale." *Measuring stress: A guide for health and social scientists* (1994).
[9] Brannon, Linda, Jess Feist, and John A. Updegraff. *Health psychology: An introduction to behavior and health*. Cengage Learning, 2013.
[10] Pollard, Lawanda J., and Larry W. Bates. "Religion and perceived stress among undergraduates during fall 2001 final examinations." *Psychological Reports* 95.3 (2004): 999-1007.
[11] Lacey, K., et al. "A prospective study of neuroendocrine and immune alterations associated with the stress of an oral academic examination among graduate students." *Psychoneuroendocrinology* 25.4 (2000): 339-356.
[12] Li, Tzai-Li, and Michael Gleeson. "The effect of single and repeated bouts of prolonged cycling and circadian variation on saliva flow rate, immunoglobulin A and α-amylase responses." *Journal of Sports Sciences* 22.11-1 (2004): 1015-1024.
[13] Walsh, N. P. "The effects of high-intensity intermittent exercise on saliva IgA, total protein and alpha-amylase." *Journal of sports sciences* 17.2 (1999): 129-134.
[14] Garrett, John Raymond, Jörgen Ekström, and Leigh C. Anderson, eds. *Neural mechanisms of salivary gland secretion*. Karger Medical and Scientific Publishers, 1999.
[15] Lupien, Sonia J., and F. Seguin. "How to measure stress in humans." *Centre for Studies in Human Stress* (2013).
[16] Herborn, Katherine A., et al. "Skin temperature reveals the intensity of acute stress." *Physiology & behavior* 152 (2015): 225-230.
[17] Hansmann, Ralf, Stella-Maria Hug, and Klaus Seeland. "Restoration and stress relief through physical activities in forests and parks." *Urban Forestry & Urban Greening* 6.4 (2007): 213-225.
[18] Hamer, Mark, Romano Endrighi, and Lydia Poole. "Physical activity, stress reduction, and mood: insight into immunological mechanisms." *Psychoneuroimmunology: Methods and protocols* (2012): 89-102.
[19] Bedimo-Rung, Ariane L., Andrew J. Mowen, and Deborah A. Cohen. "The significance of parks to physical activity and public health: a conceptual model." *American journal of preventive medicine* 28.2 (2005): 159-168.
[20] Fox, Kenneth R. "The influence of physical activity on mental well-being." *Public health nutrition* 2.3a (1999): 411-418.
[21] Retrieved from https://www.adaa.org/understanding-anxiety/related-illnesses/other-related-conditions/stress/physical-activity-reduces-st
[22] Retrieved from http://www.who.int/dietphysicalactivity/factsheet_adults/en/
[23] Southwick, Steven M., Meena Vythilingam, and Dennis S. Charney. "The psychobiology of depression and resilience to stress: implications for prevention and treatment." *Annu. Rev. Clin. Psychol.* 1 (2005): 255-291.
[24] Retrieved from https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=440116
[25] Heinrichs, Markus, et al. "Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress." *Biological psychiatry* 54.12 (2003): 1389-1398.

[26] NSF. "National sleep foundation recommends new sleep times." *National Sleep Foundation.* Retreived from https://sleepfoundation.org/press-release/national-sleep-foundation-recommends-new-sleep-times
[27] Division of Sleep Medicine at Harvard Medical School and WGBH Educational Foundation. (n.d.). Why do we sleep, anyway? *Healthy Sleep* . Retrieved from http://healthysleep.med.harvard.edu/healthy/matters/benefits-of-sleep/why-do-we-sleep.
[28] National Sleep Foundation. (n.d.). What happens when you sleep? Retrieved from http://www.sleepfoundation.org/article/how-sleep-works/what-happens-when-you-sleep.
[29] Division of Sleep Medicine at Harvard Medical School and WGBH Educational Foundation. (n.d.). Consequences of insufficient sleep. *Healthy Sleep*. Retrieved from http://healthysleep.med.harvard.edu/healthy/matters/consequences.
[30] Spira, Adam P., et al. "Self-reported sleep and β-amyloid deposition in community-dwelling older adults." *JAMA neurology* 70.12 (2013): 1537-1543.
[31] Rabbi, Mashfiqui, et al. "Passive and in-situ assessment of mental and physical well-being using mobile sensors." *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 2011.
[32] Ferreira, Denzil, Vassilis Kostakos, and Anind K. Dey. "AWARE: mobile context instrumentation framework." *Frontiers in ICT* 2 (2015): 6.
[33] Beal. "SSL – Secure Sockets Layer". http://www.webopedia.com/TERM/S/SSL.html. *IT Business edge*
[34] Everly Jr, George S., and Jeffrey M. Lating. *A clinical guide to the treatment of the human stress response*. Springer Science & Business Media, 2012.
[35] www.awareframework.com
[36] http://www.asha.org/public/hearing/Audiogram/
[37] Linn, Margaret W. "A global assessment of recent stress (GARS) scale." *The International Journal of Psychiatry in Medicine* 15.1 (1986): 47-59.
[38] Trockel, Mickey T., Michael D. Barnes, and Dennis L. Egget. "Health-related variables and academic performance among first-year college students: implications for sleep and other behaviors." *Journal of American college health* 49.3 (2000): 125-131.
[39] Harari, Gabriella M., et al. "Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods." *Computers in Human Behavior* 67 (2017): 129-138.
[40] Wang, Rui, et al. "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones." *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014.
[41] Sano, Akane, and Rosalind W. Picard. "Stress recognition using wearable sensors and mobile phones." *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013.
[42] http://www.awareframework.com/wp-content/uploads/2013/04/esm_scale_negative.png
[43] Greene, Shalom, Himanshu Thapliyal, and Allison Caban-Holt. "A Survey of Affective Computing for Stress Detection: Evaluating technologies in stress detection for better health." *IEEE Consumer Electronics Magazine* 5.4 (2016): 44-56.
[44] DeLongis, Anita, Susan Folkman, and Richard S. Lazarus. "The impact of daily stress on health and mood: psychological and social resources as mediators." *Journal of personality and social psychology* 54.3 (1988): 486.
[45] Schiffrin, Holly H., and S. Katherine Nelson. "Stressed and happy? Investigating the relationship between happiness and perceived stress." *Journal of Happiness Studies* 11.1 (2010): 33-39.
Torres, Susan J., and Caryl A. Nowson. "Relationship between stress, eating behavior, and obesity." *Nutrition* 23.11 (2007): 887-894.
[47] Weitzman, Elliot D., et al. "Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects." *The Journal of Clinical Endocrinology & Metabolism* 33.1 (1971): 14-22.
[48] Takahashi, Y., D. M. Kipnis, and W. H. Daughaday. "Growth hormone secretion during sleep." *Journal of Clinical Investigation* 47.9 (1968): 2079.
[49] Brooke, John. "SUS-A quick and dirty usability scale." *Usability evaluation in industry* 189.194 (1996): 4-7.

[50] Bangor, Aaron, Philip T. Kortum, and James T. Miller. "An empirical evaluation of the system usability scale." *Intl. Journal of Human–Computer Interaction* 24.6 (2008): 574-594.
[51] Schudy, Simeon, and Verena Utikal. "'You must not know about me'—On the willingness to share personal data." *Journal of Economic Behavior & Organization* (2017).
[22] Bosnjak, Michael, Gottfried Metzger, and Lorenz Gräf. "Understanding the willingness to participate in mobile surveys: exploring the role of utilitarian, affective, hedonic, social, self-expressive, and trust-related factors." *Social Science Computer Review* 28.3 (2010): 350-370.
[53] Glass, D. C., et al. "A telephone survey of factors affecting willingness to participate in health research surveys." *BMC public health* 15.1 (2015): 1017.
[54]
https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity
[55] http://www.awareframework.com/plugin/?package=com.aware.plugin.google.activity_recognition
[56] Larson, Reed, and Mihaly Csikszentmihalyi. "The experience sampling method." *New Directions for Methodology of Social & Behavioral Science* (1983).
[57] https://gadgets.ndtv.com/
[58] Wenz, Alexander, Annette Jäckle, and Mick P. Couper. "Willingness to use mobile technologies for data collection in a probability household panel." *Colchester: University of Essex. Unplublished manuscript* (2017).
[59] Bouida, Wahid, et al. "Willingness to participate in health research: Tunisian survey." *BMC medical ethics* 17.1 (2016): 47.
[60] Ganti, Raghu K., Fan Ye, and Hui Lei. "Mobile crowdsensing: current state and future challenges." *IEEE Communications Magazine* 49.11 (2011).
[61] https://gbiomed.kuleuven.be/english/research/50000666/50000673/cpp/ccp-dutch/onderzoek/experience-sampling-method
[62] www.pacoapp.com
[63] Cohen, Sheldon, T. Kamarck, and R. Mermelstein. "Perceived stress scale." *Measuring stress: A guide for health and social scientists* (1994).
[64] Klasnja, Predrag, et al. "Exploring privacy concerns about personal sensing." *Pervasive Computing* (2009): 176-183.
[65] Gustarini, Mattia, Katarzyna Wac, and Anind K. Dey. "Anonymous smartphone data collection: factors influencing the users' acceptance in mobile crowd sensing." *Personal and Ubiquitous Computing* 20.1 (2016): 65-82.
[66] Harari, Gabriella M., et al. "Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges." *Perspectives on Psychological Science* 11.6 (2016): 838-854.

Appendix 1

Creating an AWARE study

To create a new study campaign a few steps need to be taken. The first step is to go to the AWARE dashboard site, this allows excess of the AWARE server. However, for this master thesis, the same dashboard is created by the UT, which provides access to the university's server (this can be accessed via https://aware.ewi.utwente.nl/index.php/auth). Here an account is needed, when logged in. Figure Appendix 1.1 represents the first page on which the studies that have been created by your account can be viewed. By pressing "Create study" a new study will be created and special credentials (i.e. username and password) will be given to access the specific database for that study. Therefore, this database is only accessible by the researcher and developer.

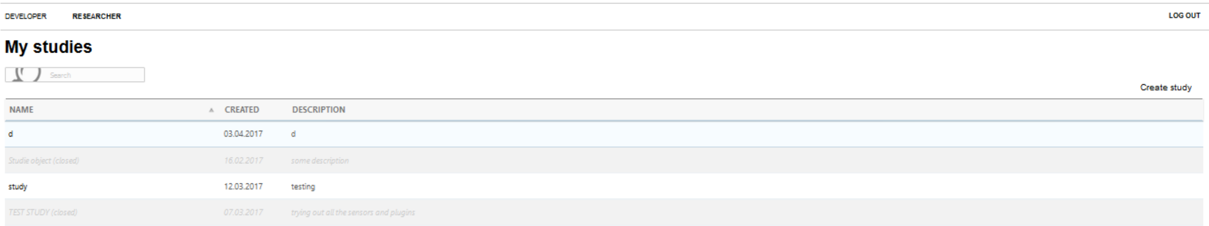


Figure Appendix 1.1: Dashboard start page

By clicking on a study a new page will be opened. This page provides all the information necessary to track and design a study. Figure Appendix 1.2 is an example of this page. The status of study indicates whether a study is open or closed. The study can be joined by smartphone users, using a specific QR-code provided on this page. The sensors can be enabled on this page as well by pressing "edit sampling". That will show all the sensors that can be activated when making a new campaign, an overview of this is shown in Figure Appendix 1.2 and 1.3.

Figure Appendix 1.2: Part one of study page

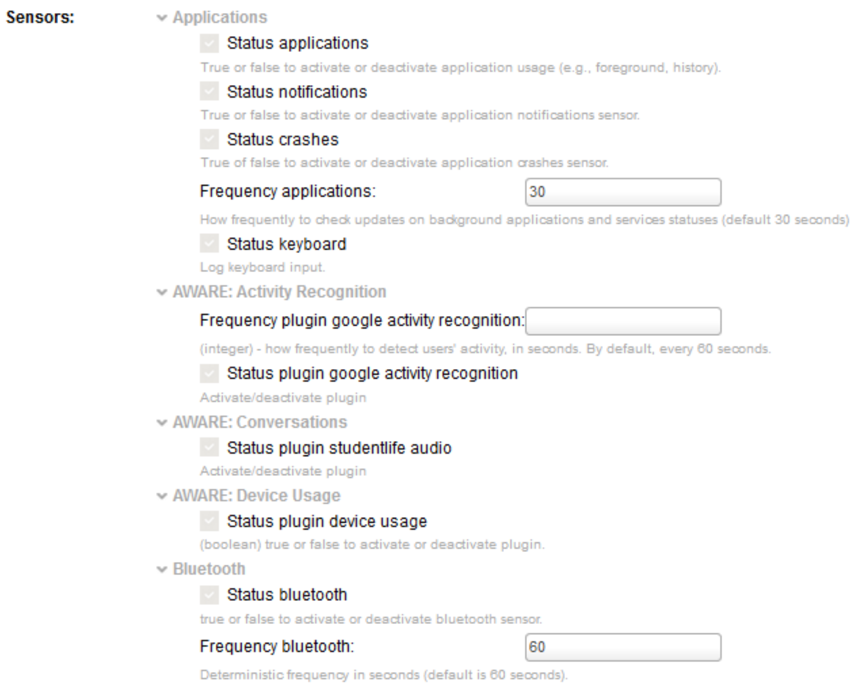


Figure Appendix 1.3: Part two of study page

When the study is created and all the sensors have been edited, the data are ready to be recorded. The calendar in Figure Appendix 1.4 shows a calendar with the number of days the data collections have been going on and it shows the recorded data. The type of records and total records are logged here as well. Lastly, in Figure Appendix 1.5 it is shown which devices are linked to the study. Prove is here provided that the ID-device is indeed a 128-bit number and does not provide any identity of the end-user. At the bottom of the page, it is possible to send ESMs to the participants of the study, by hand.

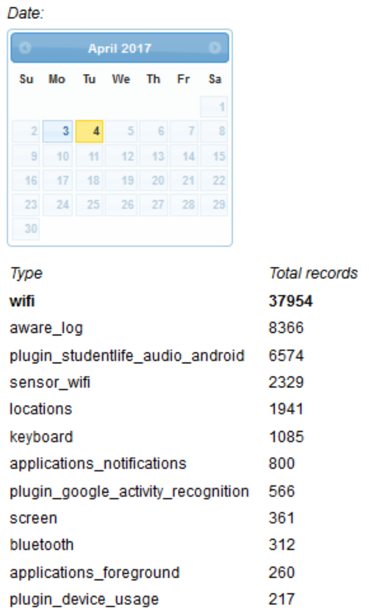


Figure Appendix 1.4: Part three study page

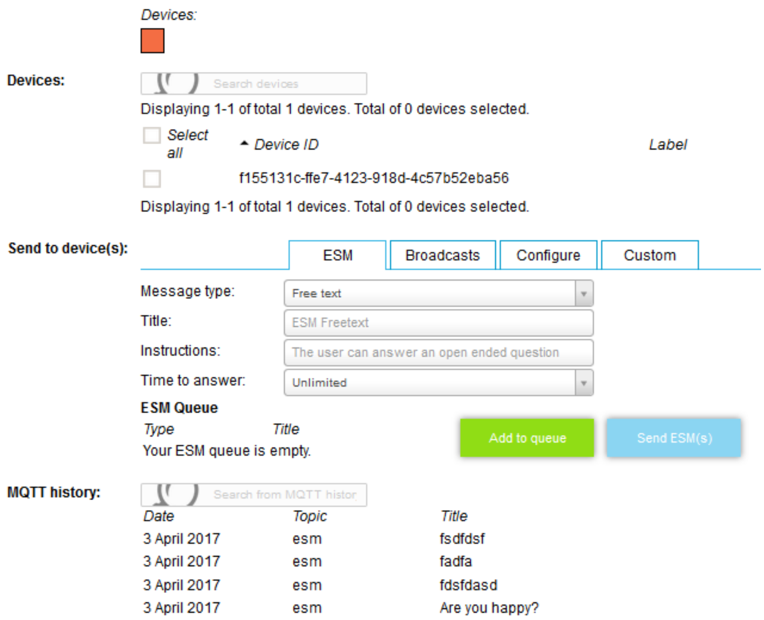


Figure Appendix 1.5: Part four study page

AWARE application

The opening page of AWARE can be seen in, Figure Appendix 1.6. It shows the UUID of the device, this is the UUID that will be recognized by the study on the UT dashboard. It also shows the sensors that are associated with this app and in the upper right corner the QR-code icon. With this icon, it is possible to connect with a study that is created on the dashboard.

When a study is joined and everything is installed it will ask permission to get access to some of the information storage of the device and provides you with messages of which snapshots are represented in Figure Appendix 1.7-1.9. This is not to get information on the content of the data but the study needs to be able to access stored data, as the data will be temporarily stored on the smartphone itself in an SQLite-database [42].

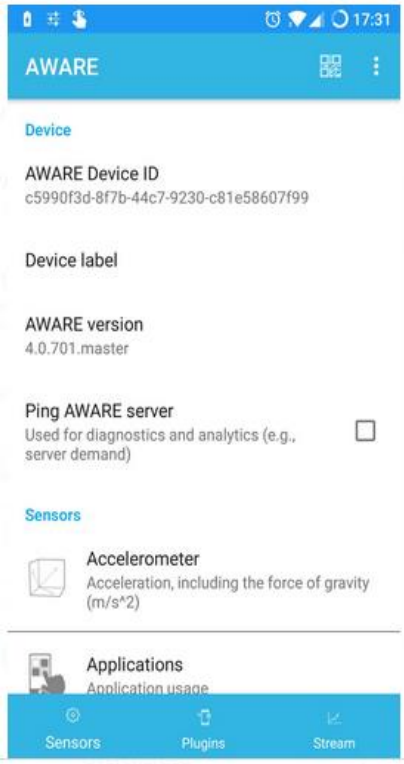
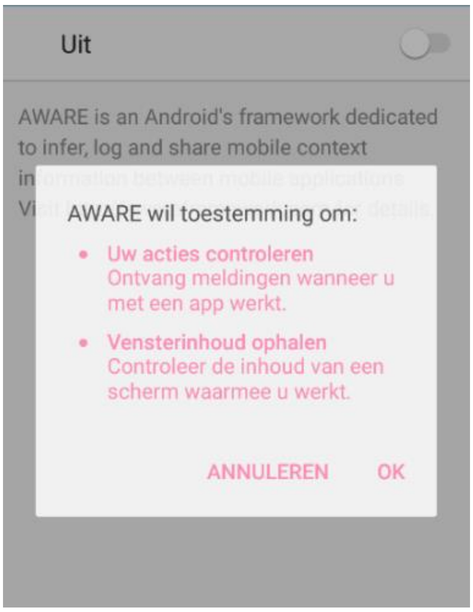


Figure Appendix 1.6: Homepage AWARE app

Figure Appendix 1.7: Overview of functions Android needs permission to access



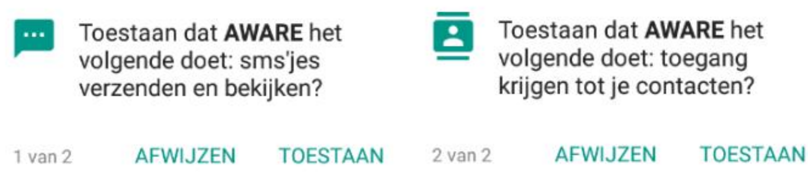


Figure Appendix 1.9: Android wants access to contacts and wants to view messages that are being received and send

Figure Appendix 1.8: Android wants permission to monitor certain activities and the content of the screen

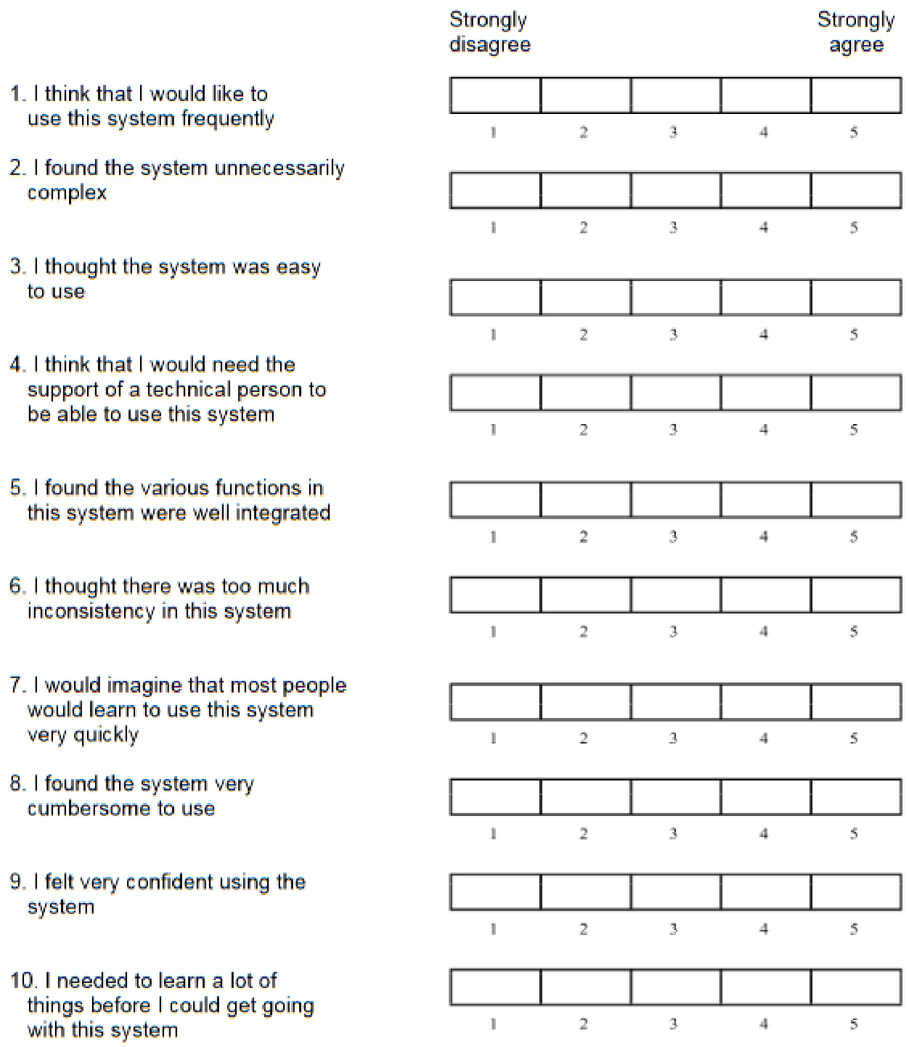
Appendix 2
Post-experiment questionnaire

Please answer the following questions and statements

Questions/statements	Answers / rating
1. I think that I would like to use this product frequently	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
2. I found the system unnecessarily complex	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
3. I thought the system was easy to use	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
4. I think that I should need the support of a technical person to be able to use this system	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
5. I found the various functions in this system were well integrated	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
6. I thought there was too much inconsistency in the system	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
7. I would imagine that most people would learn to use this system very quickly	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
8. I found the system very awkward to use	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
9. I felt very confident using the system	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
10. I needed to learn a lot of things before I could get going with this system	Strongly agree ◯ ◯ ◯ ◯ ◯ Strongly disagree
11. Overall I would rate the user-friendliness of this product as:	a. Worst imaginable b. Awful c. Poor d. OK e. Good
12. How would you rate the frequency with which questions were asked during the experiment	a. Way too much b. Sufficient c. Not enough

13. If you filled in an or c in the previous question, which frequency of answering questions would you experience as sufficient (i.e. how many times a day, you would get a notification to answer the question)?	I would think that … questions a day, I would experience as sufficient
14. How have you experienced the content of the questions?	○ Horrible ○ Sufficient ○ good
15. Please explain your answer to question 14	
16. Please provide suggestion(s) on how to improve the study (e.g. how can the study be made more pleasant)	
17. Have you experienced any technical difficulties? If so, please write them down	

Appendix 3



	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

Figure Appendix 3.1 System Usability Scale statements

GARS scale

To access stress the Global Assessment of Recent Stress (GARS) scale will be used during the *mobile stress study*. This scale is developed to assess the current perception of stress. The GARS scale is based on 8 statements that can be judged using a 10 point Likert-scale rating system of pressure, see Table Appendix 3.1. This could provide an indication of the stress perceived by someone during the day. The scale starts at None which equals zero, it indicated that someone felt comfortable (no worries and free of distress). The other end of the scale was Extreme which equals nine, it indicated that someone was very upset and might even have experience physical reactions such as headaches [37].

Table Appendix 3.1: GARS scale statements

GARS scale		
Statement number	**Statement**	**Type of answer**
1.	Please rate the next statement: Pressure related to work/job/school (whether self-imposed or not)	10 Likert scale 0 (None) – 9 (Extreme)
2.	Please rate the next statement: Pressure in interpersonal relations (family members and/or significant persons)	10 Likert scale 0 (None) – 9 (Extreme)
3.	Please rate the next statement: Pressure caused by changes in your relationships (death, birth, marriage, divorce etc.)	10 Likert scale 0 (None) – 9 (Extreme)
4.	Please rate the next statement: Pressure caused by sickness or injury (self, others, or both)	10 Likert scale 0 (None) – 9 (Extreme)
5.	Please rate the next statement: Pressure caused by financial issues	10 Likert scale 0 (None) – 9 (Extreme)
6.	Please rate the next statement: Pressure from unusual happenings (crime, natural disaster, accident, moving, etc.)	10 Likert scale 0 (None) – 9 (Extreme)
8.	Please rate the next statement: Pressure from change or lack of change in daily routine	10 Likert scale 0 (None) – 9 (Extreme)
8.	Please rate the next statement: Estimate of overall level of pressure over the past week	10 Likert scale 0 (None) – 9 (Extreme)

Appendix 4

Post-experiment questionnaire

Please answer the following questions and statements

Questions/statements	Answers / rating
○ I think that I would like to use this product frequently	Strongly agree ◯ X ◯ ◯ ◯ Strongly disagree
○ I found the system unnecessarily complex	Strongly agree ◯ ◯ ◯ X ◯ Strongly disagree
○ I thought the system was easy to use	Strongly agree ◯ X ◯ ◯ ◯ Strongly disagree
○ I think that I should need the support of a technical person to be able to use this system	Strongly agree ◯ ◯ ◯ X ◯ Strongly disagree
○ I found the various functions in this system were well integrated	Strongly agree ◯ ◯ X ◯ ◯ Strongly disagree
○ I thought there was too much inconsistency in the system	Strongly agree ◯ ◯ ◯ X ◯ Strongly disagree
○ I would imagine that most people would learn to use this system very quickly	Strongly agree ◯ X ◯ ◯ ◯ Strongly disagree
○ I found the system very awkward to use	Strongly agree ◯ ◯ ◯ X ◯ Strongly disagree
○ I felt very confident using the system	Strongly agree ◯ ◯ X ◯ ◯ Strongly disagree
○ I needed to learn a lot of things before I could get going with this system	Strongly agree ◯ ◯ ◯ X ◯ Strongly disagree
○ Overall I would rate the user-friendliness of this product as:	~~f. Worst imaginable~~ ~~g. Awful~~ ~~h. Poor~~ i. OK ~~j. Good~~
○ How would you rate the frequency with which questions were asked during the experiment	~~d. Way too much~~ e. Sufficient ~~f. Not enough~~

○ If you filled in a or c in the previous question, which frequency of answering questions would you experience as sufficient (i.e. how many times a day, you would get a notification to answer the question) ?	I would think that … questions a day, I would experience as sufficient
○ How have you experienced the content of the questions?	○ Horrible ~~○ Sufficient~~ ○ ~~good~~
○ Please explain your answer of question 14	In my opinion there is little to no relationship between the questions asked (i.e. about breakfast) and the stress is perceived throughout the day. (But this also could be a personal case, i.e. I perceive stress when someone reminds me how much stuff I need to do or when I do something with friends while I should be studying)
○ Please provide suggestion(s) on how to improve the study (e.g. how can the study be made more pleasant)	Several times I saw the notification too late, so I couldn't answer those questions anymore. Also I didn't like it to have my phone vibrating with every message I got, so a suggestion is that you could give answers to the questions several hours after you've got the notification.
○ Have you experienced any technical difficulties? If so, please write them down	I filled in the set-up survey and after that I got two notifications to fill in the set-up survey (which is already did), so I did not know if the app had saved my answers.

Appendix 5

Survey –mobile stress study

A few weeks ago I gave a presentation during one of your lectures. The presentation was about the recruitment of students for my experiment. The study was unsuccessful, due to the fact that no one volunteered to join the experiment. For this reason, I am conducting this survey, for which I would like to get your opinion on some of the aspects of the experiment. This to determine the main reasons for you not wanting to participate. This is a very important aspect of future research, as this could help future scientists to alter their study protocol. Therefore, I would like to ask you to fill in this questionnaire. I would be really grateful if you do.

Below an overview of the presentation if you need to refresh your memory.

The **aim** of the experiment is to design a stress model through behavioral stress parameters using mobile data sensing via the AWARE framework application.

Sub-objective: the designed stress model will be compared to the class schedules of the participants to see whether some schedules are more efficient than others (i.e. whether one schedule results in more stress than other schedules).

AWARE & PACO

The data that will be monitored are collected by AWARE and Paco. AWARE collects data from your smartphone using sensors and plugins. The data collection is anonymous (special encrypted ID) and stored on the UT server. Paco is a survey app, it sends you questions throughout the day. To use this app you need to log in using a Gmail account, to that Gmail account, an encrypted ID will be assigned to be able to download the data from the questions anonymously.

The next data is being monitored

AWARE

- *Physical activity* → Whether you are walking, biking, running, or in a vehicle
- *Application sensor* → Which apps you use on your phone (**ONLY NAME OF APP – NO CONTENT is measured**)
- *Device use* → Whether you use your phone or not
- *Communication sensor* → If you are calling, are being called, send a text or receive a text (**anonymously**)
- *Bluetooth* → detects surrounding devices (**anonymously**)
- *Wi-Fi sensor* → detects approximate whereabouts (**anonymously**)
- *Battery use* → detects battery levels (how much you use phone)

Paco

- *Human-based data*: survey questions → randomly send during the day

IMPORTANT INFORMATION

Some aspects of this experiment are very important. A list below lists the most important points to take into account before taking part in this experiment:

- All the data that is measured, is being measured **anonymously**
- The application sensor: it only measures **only the application names** that you have running in the foreground or in the background
- To give you anonymity and to give anonymity to your contacts a **special encrypted unique ID** is assigned to you and your contacts

- **NO content** will be measured by your apps, calls, and messages → only that something is happening
- The experiment lasts for 7 weeks, during which each day a set of two questions 6 times a day will be sent to your smartphone.

The survey consists of statements with different themes that you have to judge.

Statements/Questions	Answers
Presentation and Recruitment	
1. I found the presentation very informative	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
2. The presentation did not provide me with enough information, to want to participate in this study (i.e. the presentation should have been more detailed)	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
3. After the presentation, my opinion on this experiment was positive	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
4. Please choose from the next options. The main factor(s) for me not to want to participate in this experiment are/is:	a) The data to be monitored (physical activity, communication sensor, etc. being measured) b) The length of the experiment c) Too invasive d) Too busy (too little time) e) The questions asked during the day f) It takes to much effort to join the experiment g) No reward h) Other…
5. I would have been more motivated to join this experiment if I would have been encouraged more (i.e. have been sending more reminders to join the study, have been	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree

encouraged more to join the study via messages)	
6. Have you visited the webpage? https://mterietstap.wixsite.com/awarestressstudy	a) Yes b) No
7. Answer when 6 is answered yes. The information on the is a helpful instrument to review all the information given during the recruitment presentation.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
8. Answer when 6 is answered yes. Please rate the next statement. The information on the webpage was clear and informative.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
DATA	
9. The data to be monitored for this experiment is too invasive.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
10. I do not mind that Bluetooth is being used to detect surrounding smartphones on my phone during this experiment.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
11. I am convinced that the privacy of the people surrounded by me is guaranteed when using the Bluetooth sensor.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
12. I think the app names of my communication apps (e.g. WhatsApp, Telegram, Line etc..) should be detected during this experiment.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
13. I do not like it when the names of my social media apps (e.g. Facebook, Twitter, etc.) are being logged when I am using a social media app.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
14. It is acceptable to monitor who is contacting (anonymized) me via calls or text messages.	a) Completely disagree b) Disagree

	c) Neutral d) Agree e) Completely agree
15. It is acceptable to monitor my calls and texts, when not the content is registered but rather that the event is occurring.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
16. It is acceptable to send me questions during the day when the questions are based on scientific research.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
17. I think asking 2 questions at a time is acceptable	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
18. I think asking question 6 times a day is acceptable	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
19. Receiving the questions at fixed times throughout the day is better than receiving the questions at random times each day.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
20. I do not mind having my location tracked in order to determine whether I am stressed or not.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
21. It is different when the location is monitored via a Wi-Fi sensor than tracking location via a GPS sensor.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
22. Rank the next stress parameters, starting with the one you find least invasive to most invasive.	a) Physical activity b) Wi-Fi sensor (location) c) Device is locked or unlocked d) Calls and text received and sent e) Application name detection f) Questions asked

	g) Battery life h) Bluetooth i) Other
23. Logging and detecting smartphone data every hour of the day is too much.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
24. Logging and detecting the smartphone data for this experiment during certain parts of the day is preferable over 24/7 detection.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
SECURITY AND PRIVACY	
25. Giving the phone a special encrypted ID is save to guarantee the anonymity of the participant.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
26. The detected data on my smartphone is stored on the UT server (only accessible via a login and password). This is an acceptable and safe way to store smartphone data.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
27. I experience it as unsafe when the data collected from my phone is encrypted and then send over a Wi-Fi connection to a secure part of the UT sever	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
REWARD AND LENGTH OF STUDY	
28. A reward as a form of gratitude and motivation would not increase the chances of me wanting to join the experiment.	a) Completely disagree b) Disagree c) Neutral d) Agree e) Completely agree
Changes	
29. What change(s) would you like to see for you to reconsider joining the study?	
30. If these changes are made to the study I would reconsider joining the experiment.	a) Completely disagree b) Disagree c) Neutral d) Agree

	e) Completely agree
31. Please enter your e-mail address to enter the lottery to win one of the prices	
32. Would you like to win more prices? Enter your e-mail address here to take part in the interview.	

Appendix 6

Demographics

- What are you studying?
- What year are you in?
- What is your age?
- What is your nationality?

Presentation, installation & recruitment

- How did you perceive the information given during the presentation?
- Were there certain parts of the presentation that stood out to you?
- How was your opinion about this experiment after the presentation?
- Why did you decide not to take part in this experiment?
- Have you watched the instructional video on the web page?
    - If so, how did you perceive this video? What are your thoughts on it? (Only ask when necessary)
    - Have you tried installing the apps?
        - If so, how did you experience the installation process?
- How did you perceive a presentation as a form of recruitment?

Data

- What do you think of your physical activity being detected?
- What do you think of your smartphone use being logged?
- What do you think of your battery life being monitored?
- What do you think of your location being tracked?
- What do you think of your surrounding being checked using Bluetooth?
- What do you think of the names of the apps being logged, when you use apps on your phone?
- What do you think of your calls and text messages being logged as events?
- What do you think of two questions each time?
- What do you think of being asked questions 6 times a day?
- What would you think of the experiment if you could choose from a list of parameters to detect behavior on your smartphone?
- Do you understand the reason why no data is shown to you during this experiment?
- How would you perceive this experiment if you would be able to see the data collected on your phone?

Security and privacy

- How do you experience your smartphone being assigned an encrypted ID to ensure anonymity?
- How do you experience your contacts being assigned an encrypted ID, when logging calls and messages?
- How do you experience the data on your phone is encrypted and send over a Wi-Fi connection to the UT server?
- For the installation process of the apps, Android asks for permission to be granted access to personal parts of your phone. *Show these images*, if you would have decided to take part in this experiment. How would your thoughts have been, seeing these messages?

- These notifications are asked by Android and not by AWARE. The messages are used for other applications as well that involve sensitive data. Does this change your opinion on the previous question?

Reward & length of experiment

- How would you like to be motivated to take part in an experiment such as this?
- What do you think of the length of the experiment?

Changes

- What changes would need to be made for you to reconsider joining this experiment?