Exploring methods for self-reporting of stress

A pilot study to compare ESM on electronic devices with an online DRM questionnaire

Master thesis
at the
Department of Informatics
University of Zurich
People and Computing Lab

Prof. Dr. Elaine Huang

Prof. Dr. Thomas Fritz

First Name / Last Name: Catrin Loch
Matriculation Number: 13-718-143

Deadline: August 15, 2019

Abstract

This master's thesis explores various methods for self-reporting stress by assessing factors that influence the user experience. It proposes how self-reporting of stress should be gathered to yield a positive experience for the user. Previous work on self-reporting has not focused on the user experience, neither did it adequately evaluate what users perceive as a comfortable, non-intrusive and accurate method to collect self-reporting of stress. Results from this thesis can be used as a foundation for future research on combining self-reporting with physiological measures for assessing stress and providing immediate intervention.

Diese Masterarbeit untersucht verschiedenen Methoden zur Selbstberichterstattung von Stress durch die Erfassung von Faktoren die einen Einfluss auf das Benutzererlebnis haben. Darüber hinaus wird vorgeschlagen, wie die Selbstberichterstattung über Stress erfasst werden sollte, damit der Benutzer ein positives Erlebnis hat. Die bisherige Forschung hat sich weder auf die Benutzererfahrung noch auf eine angemessene Bewertung dessen konzentriert, was als komfortable, unaufdringliche und genaue Method zur Selbsterfassung von Stress empfunden wird. Die Ergebnisse dieser Arbeit können als Grundlage für zukünftige Forschung zur Kombination von Selbstberichterstattung mit physiologischen Messungen zur Beurteilung von Stress und zur sofortigen Intervention dessen verwendet werden.

Acknowledgments

I wish to thank the following people for their contribution to this master's thesis; Dr. Elaine Huang and Dr. Thomas Fritz, my supervisors for their patient guidance, constructive suggestions and editorial contribution; and my study participants for their participation, very valuable feedback and insights.

Finally, I wish to thank my mother and sisters, Gabriele Loch, Anne Loch and Beatrice Loch, for their continuous support and inputs; and Bruno Aegerter for his guidance during the implementation of the application for the data gathering.

Abbreviations

cf.	compare
DRM	Day Reconstruction Method
ECG	Electrocardiograph
EEG	Electroencephalogram
e.g.	For Example
ESM	Experience Sampling Method
GSA	General Somatic Activity
GSR	Galvanic Skin Response
ICD	International Classification of Diseases
JSON	JavaScript Object Notation
MARS	Mobile App Rating Scale
PPG	Photoplethysmography
PSQ	Perceived Stress Questionnaire
PSS	Perceived Stress Scale
PTSD	Posttraumatic Stress Disorder
SAM	Self-Anxiety Management
TBI	Traumatic Brain Injury
UI	User Interface
WHO	World Health Organization

List of Figures

Figure 1: Unidimensional Scaling of 28 Affect Words on Pleasure-displeasure (Horizontal Axis) and Degree of Arousal (Vertical Axis) [62: 1164] (Original Figure)....8
Figure 2: Plutchik's Three-dimensional Circumplex Model [56: 349] (Original Figure)....9
Figure 3: Stress as a Process [35: 4] (Original Figure) 10
Figure 4: Example Push Notifications of the About your Day Application on the Smartphone 24
Figure 5: First Draft iPhone Design (From Left to Right: Welcome Page, Informed Consent, Login Page, First Page Demographic Questionnaire, Second Page Demographic Questionnaire) 26
Figure 6: First Draft iPhone Survey 27
Figure 7: User Journey on the iPhone (Left: Welcome Interface, Middle: Activity Question, Right: Final Screen)....27
Figure 8: First Draft Apple Watch Survey 28
Figure 9: User Journey on the Apple Watch (Left: Welcome Interface, Middle: Activity Question, Right: Final Screen) 28
Figure 10: About your Day Logo 29
Figure 11: Smartphone and Smartwatch Questionnaire 30
Figure 12: Average Response Time per Device per Day 64
Figure 13: Average Response Time per Device 65
Figure 14: Effect of Device Type on Stress Rating per Day 71

List of Tables

Table 1: Rating of Mood Tracking Apps....20
Table 2: Emotions Reused from the Perceived Stress Questionnaire [41]....31
Table 3: Example Schedule About your Day....33
Table 4: Results of the Demographic Questionnaire....38
Table 5: Results Rank 1....42
Table 6: Results Rank 2....42
Table 7: Results Rank 3....42
Table 8: Results of the Ranking....44
Table 9: Results of the General Feedback about the Computer Questionnaire....46
Table 10: Results of the General Feedback about the Smartphone Questionnaire....48
Table 11: Results of the General Feedback about the Smartwatch Questionnaire....50
Table 12: Results Usability....51
Table 13: Results Naturalness....52
Table 14: Results about the Design....54
Table 15: Results Disruptiveness....55
Table 16: Results Environment Influence....56
Table 17: Results to In the Moment Compared to End of Day....59
Table 18: Results of the Influence of the Time of Day....60
Table 19: Results of Accuracy of the Duration of Questionnaire....60
Table 20: Results Burden....61
Table 21: Results Burden of 5 Times per Day....61
Table 22: Results Understanding of Questions....62
Table 23: Results Easiness....63
Table 24: Results of the Influence of Question on Answer....66
Table 25: Results Accuracy of Responses....67
Table 26: Results Way of Asking....68
Table 27: Results Representative Week....68
Table 28: Aggregated Results Missing a Response....70
Table 29: Results Input Type....71
Table 30: Summarized Results....72

Table of Contents

Abstract ... iii

Acknowledgments ... iv

Abbreviations ... v

List of Figures ... vi

List of Tables ... vii

Table of Contents ... viii

1. Introduction ... 1

1.1. Initial Situation and Problem Definition ... 1

1.2. Hypotheses ... 2

1.3. Approach ... 4

1.4. Goal & Contribution ... 5

1.5. Results ... 6

1.6. Structure ... 6

2. Theoretical and Methodological Foundations of Emotions ... 7

2.1. Emotion Classification ... 7

2.2. What is Stress? ... 10

2.3. Measuring Emotions and Stress ... 10

2.4. Methods for Self-reporting ... 11

2.4.1. Experience Sampling Method (ESM) ... 12

2.4.2. Day Reconstruction Method (DRM) ... 14

2.4.3. Related Research on Self-reports ... 15

3. Survey Intervention Tool ... 19

3.1. Mood Tracker and Intervention Tools ... 19

3.2. Requirements ... 21

3.2.1. Assessment Logic ... 22

3.2.2. Store Subjects' Self-reported Data ... 22

3.2.3. Push Notifications ... 23

3.2.4. Trigger Heart Rate Measures on Apple Watch ... 25

3.3. User Journey ... 25

3.4. Prototype Design ... 26

3.5. Questionnaire **29**
3.6. Implementation Difficulties **31**
4. Study Design **33**
4.1. Procedure **33**
4.2. Instrumentalization & Measures **35**
4.3. Recruiting **37**
4.4. Data Collection **38**
4.5. Analysis **39**
5. Results **41**
5.1. Comfort: Enjoyableness **41**
5.1.1. Ranking of Surveys 41
5.1.2. General Feedback Computer Questionnaire 44
5.1.3. General Feedback Smartphone Questionnaire 46
5.1.4. General Feedback Smartwatch Questionnaire 49
5.2. Comfort: Ease of Usability **51**
5.2.1. Usability 51
5.2.2. Naturalness 51
5.2.3. Design 53
5.3. Intrusiveness **55**
5.3.1. Disruptiveness 55
5.3.2. Environment Influence 56
5.3.3. Timing of Self-reports 56
5.4. Intrusiveness: Ease of Answering **60**
5.4.1. Accuracy of the Duration of Questionnaire 60
5.4.2. Burden 61
5.4.3. Understanding of Questions 62
5.4.4. Easiness 63
5.4.5. Efficiency 64
5.5. Perceived Accuracy **66**
5.5.1. Influence of Question on Answer 66
5.5.2. Accuracy of Responses 67
5.5.3. Way of Asking about Feelings 68
5.5.4. Representative Week 68
5.5.5. Missing Response 69

5.6. Effect Device Type 71
6. Discussion 72
6.1. Discussion Research Question 1 73
6.2. Discussion Research Question 2 74
6.3. Discussion Summary and Recommendations 80
7. Conclusion 83
8. Limitations 84
9. Future Research 85
References viii
Appendix A: About your Day Application xvi
Mobile Applications Review xvi
Daylio: MARS Evaluation xvii
eMoods: MARS Evaluation xx
Sanvello: MARS Evaluation xxiii
SAM: MARS Evaluation xxvi
T2 Mood Tracker: MARS Evaluation xxix
Requirement: Present Assessment Content to Subjects xxxii
About your Day: App Description xxxii
Prototype Iterations Smartphone xxxiii
Prototype Iterations Smartwatch xxxiii
Prototype Smartphone xxxiii
Prototype Smartwatch xxxv
Notification on Apple Watch xxxvi
Push Notification Schedule xxxvi
Push Notification Template Text xxxviii
About your Day: Links & Credentials xxxix
Appendix B: About your Day Study Content xxxix
Schedule for Participants xxxix
Demographic Questionnaire xl
Recruiting Flyer xli
Consent Form xli
Counterbalancing Table xliii

Participants Schedule ..... **xliii**
Examiner Schedule ..... **xlv**
Make Application Downloadable ..... **xlvi**
Application Download Description for Participant ..... **xlvii**
Computer Questionnaire ..... **xlix**
Feedback Questionnaire Computer ..... **l**
Feedback Questionnaire Smartphone ..... **li**
Feedback Questionnaire Smartwatch ..... **lii**
Overall Feedback Questionnaire ..... **liii**
Fact Sheet ..... **liv**

Appendix C: Results ..... **lvi**
Results Activity, Place and Human Environment ..... **lvi**
Response Rate ..... **lvi**
Pre-processed Results Feedback Questionnaires ..... **lvii**
Reactivity to Notification ..... **lxxviii**

Appendix E: Content of the Enclosed CD-ROM ..... **lxxviii**

1. Introduction

1.1. Initial Situation and Problem Definition

The results of the Swiss Health Survey 2012 have shown that almost one person in five often or always experiences stress at work and 18% of those in employment agree more or less with the statement that they feel emotionally consumed at work [5]. Research has shown that stress not only affects the work performance but is also associated with many physiological disorders, such as psychosis [47] or burn-outs [44]. In May 2019, the World Health Organization (WHO) emphasised that even though burn-out is added to the 11th Revision of the International Classification of Diseases (ICD-11) as an occupational phenomenon, it is not classified as a medical condition [76]. According to the ICD-11, burn-out is defined as "a syndrome conceptualized as resulting from chronic workplace stress that has not been successfully managed" [76]. Therefore, coping constructively with modern workplace stressors is in many ways a skill we have to learn [54]. First, we need to achieve consensus on the way self-reported stress should be measured. For this reason, the proposed topic, exploring methods of self-reporting stress, is considered of high interest from an academic and an applied health promotion perspective. Additionally, the topic of this master's thesis has been inspired by the work of Atz [1], Haddadi et al. [28], Hernandez et al. [33] and Van Berkel et al. [2].

Gaining an overview of the current self-reporting research has shown that the way different methods of self-reporting affect the user experience has barely been studied. Neither regarding the effect of the methodological approach on the user, nor regarding the way how the data has been gathered, i.e. by smartphone, computer or a face-to-face interview. Furthermore, research should put the user in the centre of interest, investigate the user's preference in self-reporting stress, and provide a positive experience. Hence, this master's thesis addresses the following research questions:

RQ1: How do different factors affect the user's experience of the data collection?

RQ2: How can self-reported data about stress be gathered to yield a positive experience for the user?

The thesis focuses on self-reported data about stress, because stress has been postulated to play a role in all major psychiatric illnesses and there is increasing evidence that minor life events

or daily hassles may be more powerful predictors of psychological symptoms and subjective distress than major life events [46]. The Experience Sampling Method and Day Reconstruction Method have shown to be useful tools to study subjective appraisals of stressful events that occur in the context of daily life, thus providing information on the stress-person interplay [38, 46]. Questions such as how people spend their time and how they experience the various activities and settings of their lives are also significant for researchers in diverse disciplines and are helpful to assess the current mental state of a person.

This research addresses the research questions above by evaluating two different stress self-reporting methods, the Experience Sampling Method (ESM) and the Day Reconstruction Method (DRM), and three different types of devices: computer, smartphone, and smartwatch. Furthermore, we are evaluating different factors, which might affect the user experience during the data collection and try to assess which method and input type yields a positive experience for the user. A positive experience is defined in four terms: 1) the data collection method and device comfort the user, 2) the data collection method and device are not perceived as intrusive, 3) the collected measures are perceived as being accurate and 4) the device type does not affect the stress ratings. A positive experience with the data collection method is important to reduce stress triggers caused by the actual data collection and to reduce biased self-reporting. Additionally, being repeatedly asked about particular feelings may induce those feelings or may influence users to alter their behaviour, which further affects the accuracy of the approach [48]. A nine-day study covering a within-subject design in which participants were asked to self-report their activities and feelings was conducted to address these problems and answer the research questions. A survey intervention tool with an ESM approach for iPhones and Apple Watches had been implemented and used by participants to self-report their activities and feelings during the study. The main difference between the DRM and ESM approach is that the DRM assesses the moods and feelings of participants in a reconstructive way [38]. Hence, the DRM self-reporting has been implemented with a self-hosted web survey asking participants to answer it on their computer.

1.2. Hypotheses

As described above, the thesis explores two different methods for self-reporting stress on different devices and provides insights on how self-reported stress data should be gathered to yield a positive experience for the user. Considering the related work in this field (see Chapter

2) and personal experience, the following hypotheses for the research question 2 have been defined:

1. The perceived most comfortable way to collect self-reported data about stress is with an ESM approach on a smartwatch.

   The ESM smartphone condition and the DRM computer condition are more convenient from a user interaction perspective, because the user interface of the smartwatch is very small. Nevertheless, the smartwatch is more comfortable from a personal point of view, because it reduces the workload for the user by its handiness and unobtrusiveness.
2. The perceived least intrusive way to collect self-reported data about stress is with an ESM approach on a smartwatch.

   Intuitively, responding in the moment and on a device, which is always at hand, is less intrusive compared to the ESM smartphone and DRM computer conditions. While the smartphone approach entails to locate the smartphone, which is more burdensome. The DRM computer approach is more intrusive due to the timing of the questionnaire at the end of the day.
3. The ESM is perceived as the method that results in more accurate self-reporting measures compared to the DRM method.

   The DRM method and the fact of responding at the end of the day is perceived as less accurate, because in general it is hard to remember every single moment of the day. Therefore, the ESM approach and responding in the moment is perceived as more accurate, because it is more spontaneous and covers real-time results rather than a memory.
4. The reported stress level is perceived to be unaffected by the way how the self-reported stress data is collected.

   The device type should not have any effect on the self-reported stress data, because the participants are using their personal devices. Hence, the different devices are all ubiquitous in the users' everyday live.

Within the scope of this thesis, the term *perceived* has been defined as the subjective perception of the participants in the study. An *intrusive action* is defined as an action that intrudes or interferes in one's space and results in an often-unwanted change of routine [8].

1.3. Approach

The first step entails the implementation of an ESM-based survey intervention tool on the iPhone and Apple Watch and the design of a DRM online questionnaire. The second step is a field research, which provides an overview of the physiological, psychological and behavioural processes involved in stress. This is a promising way to address how stress data should be collected as people go about their daily lives. Furthermore, this research compares the ESM, which is used to obtain self-reports about experiences at chosen moments in daily life, with the DRM, which asks participants to reconstruct their day, in a study. Participants of the study carry an Apple Watch or iPhone and get prompted with some questions about their current location, their activitiy, and who they are with, as well as information about their feelings. Subjects are also asked to fill in a DRM questionnaire on their computer. In summary, twelve participants were randomly assigned to test each of the three conditions for three days: DRM computer, ESM smartphone and ESM smartwatch. After the submission of their results on their Apple Watch, the heart rate of each participant is being measured in order to get insights on how to use this approach for further studies.

The different device types are all supplemented with a feedback questionnaire (see Chapter 4.2) to assess the efficacy of different factors on the participant's experience. The factors comfort, intrusiveness, accuracy and effect of the device type have been broken down into various subfactors to assess their perceived effect on the user experience.

- **Comfort**
  - Enjoyableness
    - Ranking of surveys
    - General feedback
  - Ease of usability
    - Usability
    - Naturalness
    - Design
- **Intrusiveness**
  - Disruptiveness
  - Environment influence
  - Timing of self-reports

        - Ease of answering
            - Accuracy of duration of questionnaire
            - Burden
            - Understanding of questions
            - Easiness
            - Efficiency
    - **Perceived accuracy**
        - Influence of question on answer
        - Accuracy of responses
        - Way of asking about feelings
        - Representative week
        - Missing response
    - **Effect device type**

The variable stress will not be induced by the experimenter; meaning a naturalistic measure of stress is achieved. The final step of this thesis is to anaylse the gathered data and to answer the research questions how different factors affect the user's experience of the data collection and how self-reported data about stress can be gathered to yield a positive experience for the user. We further provide some suggestions on how the prototype and study design can be improved and what insights these results give us on intervention tools for mitigating the stress level.

1.4. Goal & Contribution

The goal of this master's thesis is to connect the phenomenon of stress with self-reporting methods using available technologies under the condition that participants are not affected in their behaviour and feelings and by doing so to gather accurate self-assessments.

The thesis mainly contributes to the research on stress technology intervention by assessing how self-reported data of stress can be gathered to yield a positive experience for the user. This research provides a basis for designing an intervention tool for stress data gathering, which is perceived as it does not affect participants in their behaviour and therefore leads to accurate measures.

1.5. Results

The results of the field study have shown that the user experience of the data collection is not affected by the factors usability, environmental influence, understanding the questions, easiness, influence of the questions on the answer and the way of asking. However, the user experience is affected by the factors ranking, naturalness, design, disruptiveness, timing of the self-reports, accuracy of the duration of the questionnaire, burden, efficiency, perceived accuracy of the responses and the effect of the device type, which have shown to vary between the different conditions. Furthermore, the results have shown that the ESM smartwatch approach is perceived as the most comfortable and least intrusive condition. Additionally, the ESM method is perceived as more accurate than the DRM method. Finally, no difference in self-reported data across different devices could be found (see Chapter 5 & 6). Therefore, self-reported data about stress should be gathered with an ESM smartwatch approach resulting in a positive experience for the user.

1.6. Structure

This thesis is structured into nine parts. Section 1 provides an overview of the current problem and the motivation to conduct research about the problem. In order to do research about self-reporting of stress and designing a study to measure stress, Section 2 reviews the most important theoretical and methodological foundations on emotions. Section 3 outlines current available mood tracking and intervention tools for self-reporting of stress and presents the iOS survey tool, which has been implemented for the data gathering of this thesis. The study design, instrumentalization and measures to answer the hypotheses as well as the data analysis approach are presented in Section 4. Section 5 describes the results of the data collection, which are discussed and used to answer the research questions in Section 6. Section 7 provides a conclusion and is supplemented with Section 8, which focuses on the limitations of this research, and Section 9, which provides some thoughts about future research.

2. Theoretical and Methodological Foundations of Emotions

In order to minimize the number of people experiencing a burn-out, the disease should be tackled at its root cause, which is stress. The theoretical and methodological foundations of stress have been reviewed in this section to revise the basics of the phenomenon stress and how it is measured.

The first section summarizes the main theoretical approaches regarding the classification of emotions and the second section is dedicated to the definition of stress. The third section describes the different approaches to measure emotions and stress. The fourth section provides a more in-depth overview of measuring stress with self-reports and its main methodological approaches, ESM and DRM. Furthermore, this chapter is supplemented with mentionable related research on self-reports.

2.1. Emotion Classification

The study on emotion is by far not easy as over the course of the 20th century more than 90 definitions of "emotion" were proposed [56]. Since there exists little consensus on the meaning and definition of the term, it is no surprise that there is also much disagreement on the best way to conceptualize emotion and interpret its role in life [56]. Plutchik [56: 345-346] summarizes that an emotion is not simply a state of feeling, but rather "a complex chain of loosely connected events that begins with a stimulus and includes feelings, psychological changes, impulses to action and specific, goal-directed behaviour." This means, emotions do not happen in isolation but are responses to significant situations in an individual's life. Emotions are usually classified into discrete categories or into more dimensional models.

In Discrete Emotion Theory, theorists claim that there is a limited number of basic emotions and that they are distinguishable on the basis of their neural, bodily, behavioural and expressive features [13]. Additionally, these basic emotions are genetically determined products and humans across culture have the same set of basic emotions. According to a cross-cultural study by Paul Ekman and his colleagues in 1992 there exist six basic emotions [23]. The six basic emotions are anger, disgust, fear, happiness, sadness and surprise [23]. Each of these emotions can be expressed in varying degrees and each category behaves like a discrete one rather than a continuous emotional state. The activation of an emotion is said to be triggered by the brain's appraisal of a stimulus or event and other humans should be able to tell what emotion a person is feeling by looking at his or her brain activity and/or physiology [13, 23]. The critics of this approach are among others Russell and Plutchik, which claim that the Discrete Emotion Theory

cannot account for the rich variability and context-sensitivity of emotions, whereas their alternative dimensional models of emotion can [13].

The circumplex model developed by James Russell [62] suggests that emotions are distributed in a two-dimensional circular space. These two dimensions are arousal on the vertical axis from high to low and valence, meaning the continuum between pleasant and unpleasant emotions like happiness and sadness respectively, on the horizontal axis. The centre of the circle is described as a neutral valence and a medium level of arousal [61]. Emotions categorized in the left half of the circle have a negative valence, whereas emotions in the right half of the circle are associated with a positive valence. Hence, emotional states can be categorized into any level of valence and arousal. For example, excited is categorized with a high level of arousal and a high level of valence and bored can be found in the bottom left quadrant with low valence and low arousal (see Figure 1). However, the exact position of an emotion does not always correspond to our expectation on where it should be located in the two-dimensional space. This is probably due to the original data used by Russell [62] in 1980.

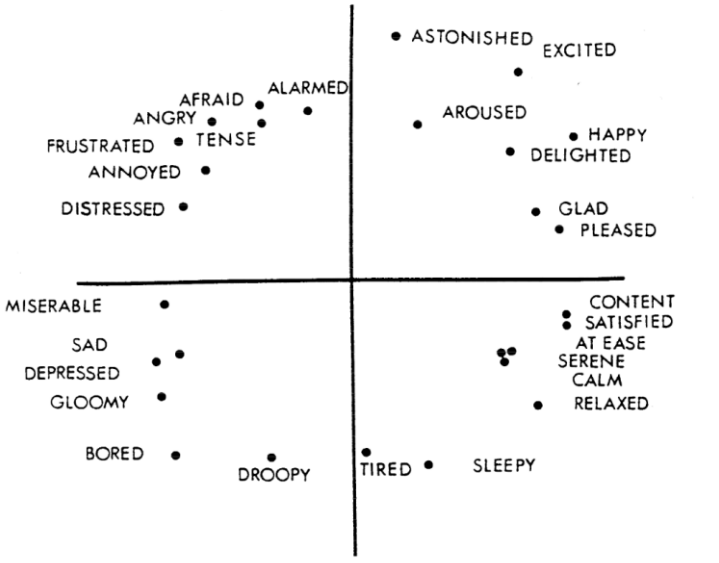


Figure 1: Unidimensional Scaling of 28 Affect Words on Pleasure-displeasure (Horizontal Axis) and Degree of Arousal (Vertical Axis) [62: 1164] (Original Figure)

The main advantage of the circumplex model is that it supports to look at different emotional states, because emotions can be demonstrated along with their relative relationship. Emotions close to each other correlate high, elements separated by an angle of 90 degrees are uncorrelated and elements separated by an angle of 180 degrees are perfectly negatively correlated. Intermediate positions in the circumplex model give corresponding intermediate emotions and

values [62]. Considering Russell's Circumplex model, from my point of view, stress would be classified somewhere between the emotions tense and depressed.

On the other hand, Robert Plutchik [56] provides a three-dimensional model that is a hybrid of both basic-complex categories and dimensional theories and founded on the Psychoevolutionary Theory. The Plutchik model arranges the eight basic emotion dimensions from the Psychoevolutionary Theory as four pairs of opposites in concentric circles, where inner circles represent more basic emotions and outer circles include more complex emotions. Similar emotions are placed close together and opposites 180 degrees apart, including complementary colours (see Figure 2). All other emotions are mixtures of the primary emotions and therefore represented with the mixture of their corresponding colour. For example, mixing rage and grief produces the mixed emotion of loathing, which is represented by mixing red and blue which results in purple (see Figure 2). The third dimension represents the intensity of emotions, which is illustrated with a cone in the circumplex model of emotions. [56]

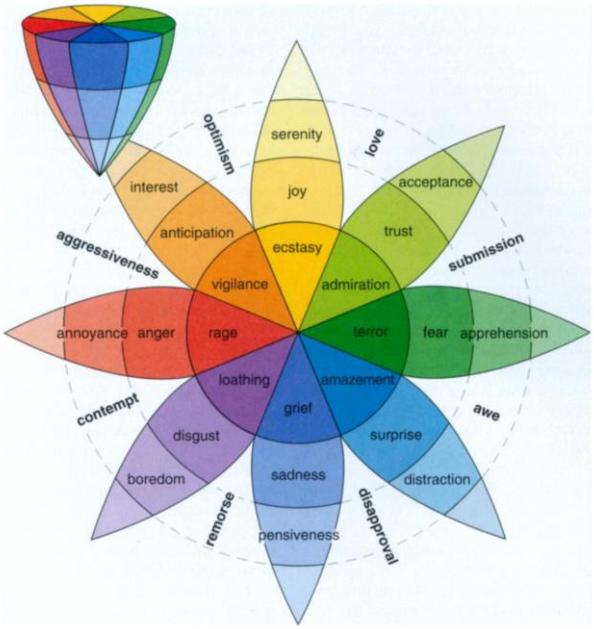


Figure 2: Plutchik's Three-dimensional Circumplex Model [56: 349] (Original Figure)

The circles in Figure 2 show the degrees of similarity among the emotions. Additionally, emotions in the blank spaces represent the emotions that are mixtures of two of the primary emotions, so called primary dyads. For example, mixing joy and acceptance produces the mixed emotion of love, which is placed in a blank space between these two emotions [56].

The similarity between Russell's and Plutchik's model is that they both use a circumplex representation, in which emotional words are plotted based on similarity [56].

2.2. What is Stress?

The meaning as well as the way in which the term "stress" has been used has not been consistent in research [35]. However, stressful events, responses, or individual appraisals of situations have always been emphasized as the central characteristics of stress [12]. For the purpose of this thesis the term "stress" is defined as a "process by which a stimulus elicits an emotional, behavioural and/or physiological response, which is conditioned by an individual's personal, biological and cultural context" [35: 4].

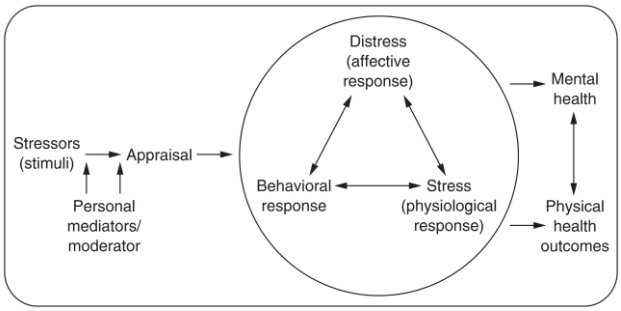


Figure 3: Stress as a Process [35: 4] (Original Figure)

The stress process includes at least three different components, which refer to the term "stress": 1) the input or stimuli, 2) processing systems, such as physiological and psychological and 3) the output or stress response (see Figure 3). The output or stress response is for example the rise in blood pressure or mental health problems. In summary, stress is the physiological response to a stimuli and appraisal, which can result in different mental health or physical health outcomes (see Figure 3). However, the reaction to this stressor is different for each individual and some events are perceived as more severe than others.

2.3. Measuring Emotions and Stress

An overview of the present research on measuring emotions has shown that there exist various approaches, such as physical measures, physiological measures, behavioural measures and self-reports.

Physical measures include body movements or postures that may have some particular meaning in terms of stress assessment [8]. Research focuses on eyelid movement, facial expressions, specific gestures, head movements, repetitive movement patterns or pupil movement [8].

Physiological measures are for example Electrocardiograph (ECG), Electroencephalogram (EEG), Galvanic Skin Response (GSR) or General Somatic Activity (GSA). The ECG measures the electrical pulse of the heart, the EEG the electrical activity produced by the brain, the GSR measures the electrical properties of the skin response to different kinds of stimuli and the GSA

measures the minute movements of the body [50]. Other measures are the respiration rate and amplitude, as well as blood pressure. These measures permit to infer mood states from the readings [50]. Various consumer wearables such as the Apple Watch and Fitbit use PPG to measure the heart rate of the user. For the study of stress, well-known indicators are skin temperature, heart rate or respiratory rate [4, 29]. Physiological sensor measures are very precise and not only used to evaluate the state of an individual, but also serve as basis for medical treatments and intervention [8].

Behavioural measures of individuals give also a detailed insight into an individual's inner self. In specific, changes in behaviours are accepted to serve as good indicators of stress effects, for example the way we perform tasks, the way we type and the way we move the mouse may signal how we feel [50]. Diagnostics of emotional states are investigated by using performance variables such as time to complete tasks and error rates such as the number of typos [50].

The most known self-reporting mechanisms are questionnaires, which are an inexpensive approach to collect vast amounts of information [8]. These self-reports give among others the possibility to measure anxiety, depression, negative mood states and daily stressors. Self- reports may be the only way to access experiences and internal states, such as happiness. However, many researchers remark the inaccuracy of self-reports. Some widely used and validated instruments include the 30-question "Perceived Stress Questionnaire" by Levenstein et al. [41] and the "Perceived Stress Scale" by Cohen et al. [11]. Self-reports assess traits of negative affectivity, which is associated to an elevated cortisol level [22]. In order to improve the inaccuracy of self-reports, many researchers make use of this fact and combine their self-reports collection with saliva sampling [22]. This is just one of many approaches to improve the inaccuracy of self-reports by adding a physical or physiological measure.

2.4. Methods for Self-reporting

In the last few years the generic name of *everyday experience methods* has matured from the status of being a promising innovation to a standard tool of self-reporting used in social-personality psychology [59]. This is also reflected by the number of published studies, which use an everyday experience method approach. One major component which supported the rise in popularity is the technological progress. It does not only support the collecting possibilities of self-reports, but also the availability of statistical tools. The term *everyday experience methods* includes methodological approaches like the Experience Sampling Method, Ecological Momentary Assessment, Ambulatory Assessment or Diary Methods [59]. These terms are

mostly used interchangeable and consolidated in the term Intensive Longitudinal Methods [73]. The Experience Sampling Method and Day Reconstruction Method are covered in more detail below.

2.4.1. Experience Sampling Method (ESM)

The most similar term to ESM is EMA, Ecological Momentary Assessment, which is often used interchangeable and was developed by Shiffman, Stone and colleagues [66, 68]. EMA is related to healthcare and has its origins in the behavioural medicine field. EMA measures the quality of people's lives by repeated sampling of subjects' current behaviours and experiences in real time in their natural environment [66]. The multiple assessments over time, provide a picture of how the subjects' experiences and behaviours vary over time and across situations. This approach aims to minimize recall bias, allows to study microprocesses that influence behaviours in real-world contexts and to maximize ecological validity [66]. EMA studies make use of different technologies ranging from written diaries and telephones to electronic diaries and physiological sensors. "In summary, EMA aims to assess the ebb and flow of experiences and behaviour over time, capturing life as it is lived, moment to moment, hour to hour, day to day, as a way of faithfully characterizing individuals and of capturing the dynamics of experience and behaviour over time and across settings." [66: 5] Combining EMA with the rapid technological advances has resulted in Ecological momentary interventions (EMI) [49]. These adopt mobile devices for the delivery of treatments in the daily life of patients, while they are engaging in their daily life activities [49].

In the 1970's, Csikszentmihalyi and colleagues [14, 16] introduced the Experience Sampling Method (ESM), which refers to capturing individuals self-reported experiences (thoughts, states, events) using a random sampling approach. It focuses on measuring the subjective state of individuals, for example happiness, stress, pain and many other emotions and feelings in one moment [3, 66]. Participants in ESM studies are prompted on their electronic beeper to record where they are, what they are doing, and how they feel at preselected but randomized time points throughout the day [19]. This technique provides a rich description of a sample of moments in respondents' lives combined with the precision of scaled questionnaire measures, while avoiding the distortions that affect the delayed recall and evaluation of experiences [31]. Therefore, its applicability and related methods have been demonstrated in various studies, for example covering psychiatric symptoms like schizophrenia [19], young delinquents [16] or

substance use [65]. This has also led to the fact that the terms EMA and ESM cannot be clearly distinguished anymore.

In summary, the ESM measures the quality of life and evidence points towards greater efficacy when ecological momentary interventions (EMI) are integrated with real-life assessments using the experience sampling methodology [49]. Finally, previous studies have demonstrated the feasibility, validity and reliability of ESM and EMI in psychiatric populations [47].

2.4.1.1. Advantages of ESM

One of the main advantages of ESM is its immediateness and that is asks about the experience and context [1, 31]. Additionally, it allows one to study a phenomenon in the real world rather than in an artificial environment, which implies a high ecological validity [1, 46, 52]. Comparing results in ESM allows for intra-personal as well as inter-personal comparison [1]. ESM also gives the possibility to beep subjects at random moments [46, 52]. Moreover, it allows to assess a detailed record of an experience including multiple assessments of context and constituting an excellent tool to study the interaction with contextual features [46, 52]. Therefore, it is very accurate and includes barely any recall and memory bias [3, 46]. Furthermore, the longitudinal character of the data allows to investigate variation over time and unconscious processes may be made explicit in the data [46]. This is further supported by the fact that new statistical approaches, such as multi-level and mixed-effects models, allow analysing the unbalanced hierarchic ESM datasets with variation at the level of the subject and the level of the moment [1].

2.4.1.2. Disadvantages of ESM

A small number of studies have shown how sampling frequency in ESM influences the willingness to take part and the compliance. Consequently, the main disadvantage of ESM is that it is usually highly obtrusive, and this leads to self-selection, attrition and biased data [1, 31]. Furthermore, it involves high levels of participant burden and provides little information about uncommon or brief events, which are rarely sampled [31]. In general, ESM studies are also very time-consuming and demanding for participants. Myin-Germeys [46] have further stated that although the ESM has a longitudinal structure, most effects will not last over beeps and therefore still require cross-sectional analyses. One way to adapt the design to the temporal dynamics of the processes of interest is to shorten the time intervals between beeps [46]. However, this may also increase the intrusiveness of the method and possibly result in induced rather than recorded experiences [1, 46]. Hence, it is highly recommended that researchers

carefully construct the ESM questionnaire and balance the number of reports to avoid reactivity to the method [1].

2.4.2. Day Reconstruction Method (DRM)

The Day Reconstruction Method (DRM) was first presented by Kahneman et al. [38] and refers to having participants fill out surveys about one's life the next day in diaries. This method represents a retrospective survey approach to collect data describing personal experiences on a given day, through a systematic reconstruction conducted on the following day [37]. It combines the strengths of time-budget measurements [36, 60] and experience sampling [69] and further employs techniques grounded in cognitive science. Time-use studies collect so- called time-budget measurements by determining how people allocate their time during an average day [36]. The results of Kahneman et al. [38], indicate a close correspondence between the DRM and established results from experience sampling.

The DRM is designed such that it asks respondents to reconstruct the previous day by completing a structured self-administered questionnaire, which is designed to reduce bias [38]. Respondents first revive memories of the previous day by constructing a diary consisting of a sequence of episodes. Afterwards, they describe each episode by answering questions about the situation and about the feeling that they experienced [37].

The goal of the DRM is to provide an accurate picture of the experience associated with activities, for example commuting and its circumstances. Regarding Kahneman et al. [38], the ESM is the gold standard to which DRM results must be compared to.

2.4.2.1. Advantages of DRM

Kahneman et al. [37] have mentioned multiple advantages the DRM brings with it. The main advantage of the DRM approach is that it allows for a joint assessment of activities and subjective experiences. It further provides information about the duration of each experience and allows for duration weighted analyses of experiences. Compared to other experience sampling methods it comes with a lower respondent burden and includes a more complete coverage of the day than typical for experience sampling methods. Additionally, researchers are given a high flexibility in adapting the content of the instrument to the needs of the specific study and a lower susceptibility to retrospective reporting biases than typical for global reports of daily experiences. Compared to experience sampling it also provides time-budget information, which is not collected effectively by experience sampling. Another advantage of

the DRM approach is that it supports both between-subjects comparisons (e.g. different age groups) and within-subject comparisons (e.g. different situations or times of day).

2.4.2.2. Disadvantages of DRM

The main disadvantage of the DRM is the convincing evidence that most people suffer from recall bias and use heuristics to fill in the self-reports [1, 3]. Additionally, compared to the ESM, the DRM does also place a substantial demand on the participant with the burden of repeated queries and responses [3]. Further research has also shown that respondents tend to rely on a "peak-end" rule when reflecting on their pain or mental health issue [58]. Participants give more weight to the peak levels and to the most recent levels of experience rather than equally weighting each instance [58]. Moreover, Stone et al. [67] have analysed whether retrospective questionnaires or experience sampling assessments produce similar results for an individual's self-reports. The authors have found very poor correspondence between the two measures and that patients with mental illness are more likely to recall emotionally negative than positive feelings, which leads to under- or over-reporting certain measures [67].

2.4.3. Related Research on Self-reports

This section summarizes the related results of three research questions regarding self-reports, which have shown to be of high importance for the study design. These questions are: 1) are paper-based and electronic self-reports equivalent, 2) how can self-assessments and smart devices be combined and 3) can results reported with ESM and DRM be compared.

2.4.3.1. Self-reports: Paper-based or Electronic?

Gwaltney et al. [27] have shown that simply moving a questionnaire from paper to a screen-based electronic device should not be problematic. They have analysed 65 studies that directly assessed the equivalence of computer versus paper versions of self-assessment used in clinical trials. Their results have shown that paper-based and screen-based questionnaires were equivalent. Nevertheless, further evidence is still needed to demonstrate that scores derived from a computerized measure do not differ from scores derived from the paper and pencil version [27]. Reasons for the potential differences between these measures are the way of how items and responses are presented and that some individuals may have difficulties interacting with computers [27]. However, people nowadays are much more used to carry around a technical device, which implies a lot less intrusion than using paper and pencil [1]. Additionally, missing data within a computerized assessment can be reduced by requiring completion of an item before the patient can move on to the subsequent question [27]. Another advantage of

computerized assessments over paper and pencil assessments is that computerized assessments can handle complex skip patterns and reduce the effort and error involved in entering paper data [27]. Implementing a diary study in an electronic format provides the possibility to implement sophisticated designs to ensure valid representation of the patient's experience [27]. Additionally, computerized assessments address one of the main problems of paper-based questionnaires, they can capture time-tag records to document timely compliance and can increase general compliance [27].

2.4.3.2. Self-assessments on Smart Devices

Smartphones are the permanent companion of people all over the world and have marked their position as pervasive technology. In addition, extensive efforts have been made in both academia and industry to develop smart wearables for health monitoring and sensing [9, 28]. Overall ubiquitous mobile and wearable devices can support the strategies to target mental health and well-being by having the knowledge about the user's current affective and emotional state [2, 28]. In particular, wearables provide the possibility to gather even richer and more detailed insights at various levels of causality (biological, psychological, social) [48]. This knowledge can further be used to identify critical states and is the basis for new interventions, such as targeted therapies and prevention through early detection [28].

Combining ESM with mobile devices has been used by scientists from various disciplines to gather insights into the intra-psychic elements of human life [2, 55]. For example, to assess the temporal relationship of daily life stress and bowel symptoms [10] or to collect music listening data to advance the accurate measurement of everyday, personal music listening [57]. Van Berkel et al. [2] give an overview of studies which use an ESM approach on mobile devices and provide researchers with practical advice on running ESM studies on mobile devices. According to them, the main advantages of a smartphone-based ESM study are improved data quality through validation, context reconstruction, real-time study status, advanced question logic and rich media collection [2]. A more specific literature review has shown that there do not exist many smartphone-based ESM studies to assess stress [1, 25, 74]. Furthermore, that the type of device used to deliver an intervention as well as specific design choices may be associated with differences in how the intervention content is used [45, 26]. Many researchers currently focus on developing stress interventions tools which provide the user with an intervention mechanism right in that stressful moment. For example, a message to cheer up, a breathing exercise, mobile gaming or acupressure [43, 53, 54, 75]. These tools all try to raise

the awareness and provide feedback in a stressful situation, but do not consider the user's preferences and how the method used to gather the self-reports impacts their perceived stress level.

2.4.3.3. Comparing Self-reported Results of ESM and DRM

Comparing self-reported results of ESM and DRM is of interest, because DRM data is based on recall and depends on memory, whereas ESM data is gathered in the moment. Therefore, researchers are trying to determine whether these two methods of data gathering lead similar results, despite the dependence on memory in the DRM. As described in chapter 2.4.2, the DRM has been introduced to overcome the disadvantages of the ESM, such as the high level of participant burden or the high cost [38]. This results in the fact that the DRM has its roots in the ESM, and it is also intended to reproduce the information that would be collected by probing experiences in real time [38]. Hence, as mentioned above, Kahneman et al. [38] explain that DRM results must be compared to the gold standard which is the ESM. The authors have conducted a DRM study with 909 working women and compared these results with established ESM data with similar affect categories, to verify their claim [38]. Their analysis has shown that the results of the DRM data collection can be compared to the ESM data set [38]. For example, the "tiredness" ratings resulted in a V-shape for both datasets and the correlations of affect were high for positive emotions (0.7) and negative emotions (0.4) [38]. However, the main limitation of this study is that the DRM and ESM ratings were not reported by the same sample and not at the same moments in time.

In 2010, Dockray et al. [21] examined 94 working women for two days to validate the DRM by comparing these reports to six ESM ratings. The authors were able to compare the average DRM episode rating of a certain hour to the ESM ratings of approximately that time, because participants could rate their mood in several affect categories which were the same for both methods. Their results have shown that the correlations between ESM and DRM ratings vary between 0.52 and 0.79. In other words, the intensity and variation of affect are similar for ESM and DRM and consequently the DRM can be used to overcome the disadvantages of the ESM [21]. Atz [1], Bylsam et al. [7] and Hedges et al. [30] have each conducted similar studies and their findings suggest that the two methods will yield similar estimates for individuals over time. However, the limitation of these kind of studies is that ESM and DRM reports were provided to the participants for the same time interval [20, 32]. Hence, it is likely that early ESM mood reports influence the later that day collected DRM report [20, 32]. For example, the

accuracy of the DRM report is increased due to the fact that participants pay more attention to their emotions and might remember some ratings while filling in ESM self-assessments during the day [20, 32]. The influence of participants giving both ESM and DRM ratings concerning the same day is also called carry-over effect [20, 32].

Overall, the literature research has shown that there are only a few studies that investigate the association between ESM and DRM results [1, 21, 38, 39, 40, 72]. They have found that DRM ratings can be compared to ESM ratings when they use the same set of emotions and especially happiness ratings of the two methods are highly correlated. In the end, it is uncertain to what degree the DRM actually eliminates recall biases, reduces participant burden or increases compliance compared to current experience-sampling methods that employ modern technologies [20]. Additionally, it is still unclear to what extent the ESM and DRM ratings disperse.

3. Survey Intervention Tool

Chapter 2 has given an overview of the theoretical and methodological foundations of emotions, which have been considered in the development of stress intervention tools. For example, applications for the collection of stress data or tools which provide the user with real-time intervention in a stressful situation. This section summaries the main components of the iOS survey intervention tool "About your Day", which has been designed and implemented for the data-gathering of this thesis. The first section provides an overview of a selection of mood trackers and intervention tools, which already exist and have therefore been used as basic approach to define the requirements of our application. The second section briefly describes the functional requirements of the survey tool and how these components have been implemented in iOS. The third section entails the user journey of the tool and the fourth section describes the prototype design. In the fifth section the questionnaire, which represents the content of the tool, is being specified. The final section summarizes the implementation difficulties which occurred during the whole implementation process.

3.1. Mood Tracker and Intervention Tools

An overview of current available mood tracking and intervention tools is used as baseline and inspiration for the implementation of the survey tool of this thesis. Five free available mood tracking and intervention applications have been selected considering their reputation, i.e. the number of times they have been listed in an online review for good mood tracking apps (see Appendix A). The selected applications have then been rated with the Mobile App Rating scale for health mobile apps [70]. The Mobile App Rating scale (MARS) consists of five broad categories of criteria: engagement, functionality, aesthetics, information quality and subjective app quality, which are assessed using a 23-item questionnaire (see Appendix A). The criteria engagement is described with the factor's fun, interesting, customisable, interactive (e.g. sends alerts, messages, reminders, feedback, enables sharing) and well-targeted to audience [70]. The functionality includes the aspects: app functioning, easy to learn, navigation, flow logic and gestural design of app [70]. The aesthetics are being assessed with the elements: graphic design, overall visual appeal, colour scheme and stylistic consistency [70]. The information score is measured by determining whether the application contains high quality information (e.g. text, feedback, measures, references) from a credible source [70]. The rating results by applying the MARS questionnaire for each selected application (see Appendix A) are summarized in Table 1 and used to give an overall assessment of the quality of each application. An in-depth analysis

or discussion of the ratings of each application was not done, because the goal of the analysis was to gain an overview of available applications and collect some key takeaways, which can be integrated into the implementation of the "About your Day" application.

Logo	App	Engagement Score	Functionality Score	Aesthetics Score	Information Score	Quality Score	Subjective Quality Score
	Daylio	3.6	5.0	4.3	3.7*	**4.1**	**3**
	eMoods	3.0	4.25	4.3	3.7*	**3.8**	**1.75**
	Sanvello (Pacifica)	4.8	4.5	4.3	4.0*	**4.4**	**4**
	Self-Anxiety Management	2.8	4.25	4.0	4.0*	**3.8**	**2**
	T2 Mood Tracker	3.2	4.5	3.7	4.3*	**3.9**	**1.75**

Table 1: Rating of Mood Tracking Apps
*Remark: not all sub factors of the information score could be assessed for these applications

Daylio [17]

Daylio is a mood-tracking app that gives the user the possibility to identify and track his feelings without typing any word. It provides several videos that portray different moods from which the user can choose the one which reflects his mood and feelings the best in that moment. Additionally, moods can be matched to activities and additional notes can be added.

eMoods Bipolar Mood Tracker [42]

eMoods Bipolar Mood Tracker is not just for people who have bipolar disorder it can also be useful for people with general anxiety or depression. Users have the possibility to rate their mood, specifically irritability and anxiety, as well as to track some behavioural data like the hours of sleep or current medications.

Sanvello [63]

The idea of Sanvello is to track daily activities by writing down or using audio recording based on cognitive behavioural therapy combined with relaxation and wellness techniques. The app helps then to find out what triggers' different moods and emotions, especially stress and anxiety.

Self-Anxiety Management (SAM) [64]

Self-Help for Anxiety Management provides the user with information, helpful resources, guidance and 25 self-help tools to manage their anxiety. It helps to figure out what's making

you anxious or stressed while also suggesting ways to combat it by providing the possibility to connect with other people combating their anxiety through a closed social network.

T2 Mood Tracker [71]

The T2 mood tracker tool helps to track and assess anxiety, depression, overall well-being, Traumatic Brain Injury (TBI), stress, Posttraumatic Stress Disorder (PTSD) or any customized emotional category. Users can rate their emotions with sliders which are automatically displayed in a graph. The application supports to discover patterns in users' emotions or to evaluate possible triggers of negative emotions.

Main Takeaways

After analysing the five applications above, this paragraph summarizes the main takeaways, which need to be considered for the design and implementation of the "About your Day" application. First, a neat design can highly improve the usability and efficiency of an application. For example, too many options and possible interactions are most of the time overwhelming for the user and the whole application seems to be overloaded. Second, the more clearly structured, coherent and consistent the design is the more intuitive is the whole user experience and the more is the user willing to use the application again. Third, to engage the user with the application it needs to be self-explanatory. Otherwise the user does not use it anymore, because he needs to put too much effort into figuring out how it works. Fourth, the possibilities of interaction always need to consider the purpose of the application. For the "About your Day" application this means, that the focus should be to gather self-reported data. Therefore, the purpose should also be reflected in the number of functionalities the application entails. Fifth, the colours in the application should be coordinated throughout the whole application, considering the target user group and the meaning they implicate for different cultures.

3.2. Requirements

The requirements of the iOS intervention tool are deduced from the goal of this thesis and other intervention tools (see Chapter 3.1). The application needs to cover the following general requirements and functionalities:

- Present assessment content to subjects
- Manage assessment logic such as random order of questions
- Store subjects' self-reported data

- Ability to prompt subjects to complete assessments (see Section 3.2.3)
- Manage prompting schedules (see Section 3.2.3)
- Trigger heart rate measures on Apple Watch

Some of the requirements above, which are not self-explanatory, are specified in the following. All additional information about the survey intervention tool can be found in the appendix (see Appendix A).

3.2.1. Assessment Logic

The tool has been implemented such that the first three questions about the activity, place and human environment are always displayed on the first three screens of the assessment. All questions about the feelings have been implemented such that those questions are presented in a random order every time the participant is asked to fill in an assessment. This random order helps to overcome a possible influence of the order of the feeling questions on the participants' answers. The random order is defined with a random array, which is generated once a participant opens the application. For each of the eleven feeling questions a number has been assigned and the function *generateRandomQuestionArray* goes through the list of numbers and generates a random order. This random order is saved in the array *numberArray* and when the user has answered the question about the human environment, the interface calls the class *QuestionPhone* or *Question* (for the Apple Watch) with the function *allQuestions*. This function converts the questions saved in the JavaScipt Object Notation (JSON) such that these are accessible and can be used to replace the default content. If the content has been converted, the *viewDidLoad* function loads the assigned content, such as the question text and labels, which has been assigned to the number in the random *numberArray* and overwrites the default content with the assigned content on each screen.

3.2.2. Store Subjects' Self-reported Data

In the setup of the application, each time a participant opens the questionnaire a struct is created. A struct is a custom data type, which provides storage of data using properties with extended functionality. Inside the braces of a struct the properties of this specific struct are defined and they are like attributes of the struct. The defined struct in the programmed application is called *passableData*. After passing the first screen the struct is activated and the *user*, *numberArray* and *currentDate* are saved. The struct is then passed on to the next screen, where the selected response of the participant is saved into the struct. For each response an empty space has been defined in the struct and each time the user selects an answer, the response is saved and struct

is passed onto the next screen. The final struct, which includes all answers, such as the *user*, *currentDate* and *endDate*, is converted into a JSON on the submit screen and passed to the server by clicking the "Submit" button. Multiple researchers have mentioned that fake compliance is a big problem in self-reporting data (see Section 2.4.3.1). Our approach to overcome this problem was to attach the timestamp to the user with each data entry. When the user starts the survey, the timestamp is gathered as starting time and when the user finishes the survey and submits his answers, the finish timestamp is saved. Additionally, we made use of the fact that typical Apple devices are named after their owner and therefore accessed the name of each device. We saved this information in the *user* field to identify the responders and to complement the timestamp information. These two implementations help to avoid fake compliance and to retrieve the time it took the participants to fill in the survey.

3.2.3. Push Notifications

The ability to prompt subjects to complete assessments and manage a prompting schedule has been implemented with push notifications. A prerequisite for having users completing a survey is that they need to get notified at random times during the day. One Signal is a notification platform on which users are added as subscribers and push notifications can be managed individually [51]. The One Signal Platform further provides management of prompting schedules on its dashboard [51]. To make use of the One Signal service, their package needs to be included in the application. After downloading the application, each iPhone is automatically registered on the One Signal Platform. Therefore, each device could be linked to its owner and the notification schedule could be managed individually according to their assigned schedule. As mentioned above, the One Signal platform further provides the possibility to manage the push notification schedule. The administrator can create a new push, select the audience to which the notification should be sent to and add the content of the notification. In fourth section of the new message interface on OneSignal, the administrator has then the possibility to schedule the delivery or send the notification immediately. The platform has also many other features such as to send additional data fields, attach action buttons or optimize the delivery by user time zone. In order to minimize the workload during the study, the notifications during the About your Day study have been scheduled some days in advance. On the One Signal website under delivery and scheduled messages, all the scheduled messages were then available and could be adjusted and checked if necessary.

An example notification is shown in Figure 4. The notification is key to open the application on a device, Apple Watch or iPhone. Regarding the Apple Watch, a user needs to click into the notification on his Apple Watch (see Appendix A), then the About your Day application is opened and the user is able to fill in the survey. If a user swipes right or clicks into the notification on his iPhone (see Figure 4) and selects open, again the About your Day application opens and the user is able to fill in the survey. Without receiving a notification, the user is not able to fill in a survey on his iPhone. Due to technical difficulties this could not be implemented on the Apple Watch. Hence, the user has always access to complete a survey on his Apple Watch.

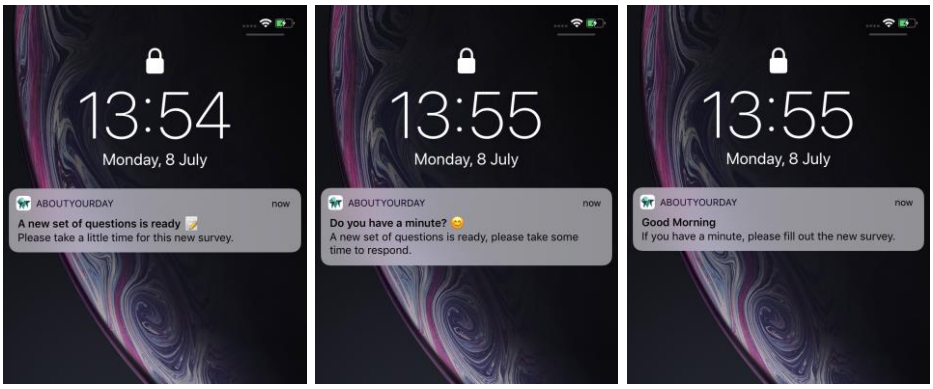


Figure 4: Example Push Notifications of the About your Day Application on the Smartphone

In order to not bore the user too much, five different notification texts have been created, such that the user received a notification type just once a day (see Appendix A). Additionally, the notification did not expire, because we tried to find out whether the number of times the participants were prompted had an effect on the user experience. The notification content has been made more interesting by adding time specific content, such as a title saying "Good Morning", or by adding an emoji. Additionally, the text has been kept short, in order to communicate immediately what the user had to do. The notifications of the application are visibly not different from any other notification, including the name and logo of the application, the time the notification has been received, a title and a short text (see Figure 4).

Many researchers [45] implemented different frequencies, such as daily between 5pm and 8pm or twice a week between 5pm and 8pm, to assess the impact of the frequency on the self-reported answers. Atz [1] has shown that the reactivity has no substantial effect and proposes that researchers should minimize the complexity of the input, i.e. the amount and type of question, interface and integration, rather than reducing the number of prompts. Therefore, the frequency of the iPhone and Apple Watch questionnaires has been limited to 5 times per day, which also includes sending five notifications per day to the user. As a control possibility, the

One Signal platform also provides a dashboard, which covers information about the status of the message, date and the number of users has clicked the notification.

3.2.4. Trigger Heart Rate Measures on Apple Watch

The Apple Watch uses two green light-emitting diodes (LEDs) paired with light-sensitive photodiodes to measure hear rate by detecting the amount of blood flowing through the wrist at any given moment [50]. During a heart beat the blood flow is greater and therefore the green light absorption is bigger. Between heart beats, the blood flow is less and so is the green light absorption. The device can then calculate the user's heart rate by flashing its LEDs hundreds of time per second [50].

The heart rate of users is gathered by triggering an activity on their Apple Watch. This part of the application has been implemented by Jan Gugler within the scope of his independent studies at the University of Zurich. After the user pressed the submit button on his Apple Watch the Watch Kit application triggers an activity and gathers the current heart rate of the user. The heart rate is then be measured for 10 minutes every 5 seconds. The data is saved in the format of a JSON which is further passed to the server of the About your Day study (see Appendix A). Due to technical reasons, the final screen of the Watch application sometimes showed up multiple times, while the heart rate data has been gathered. However, participants were informed about this issue in advance and therefore knew how to handle it.

3.3. User Journey

The following section describes the user journey on the iPhone and Apple Watch. The participant starts off with a push notification informing that a new survey is available. After opening the notification (see Section 3.2.3), a welcome screen is displayed showing a short text saying that a new survey is available and the button "New Survey". The user then clicks on this button and is lead to the screen with the activity question. The proband has 10 possible answers to choose from. After selecting, an answer the second question about the place is displayed. Once again, the user has 10 possible answers to select from. After selecting, the human environment question is displayed. The user has to choose from 4 different answers and again arrives at the next screen after selecting one. After those three questions, the feeling questions are being displayed which means that 11 feelings with the same answer options are displayed. Again, each time the user selects an answer the next question appears. After those 14 questions the user is then asked to submit his answers by clicking "Submit". If the submission has been successful, the proband arrives at the final page stating that the answers have been successfully

submitted and the participant receives a new notification if a new survey is available. If the submission has not been successful, the interface displays a short message saying that the answers could not have been saved and the user should try it again.

If a participant opens the iPhone application without receiving a notification, a default interface informs the user that he will get a notification if a new survey is available and that he should exit the application by clicking “Exit”. If a user opens the Apple Watch application without receiving a notification, he is able to complete a survey and the journey takes place as described above.

3.4. Prototype Design

For the design of this prototype, we guided ourselves considering the different currently existing intervention tools (see Chapter 3.1) and suggestions supported by research (see Section 2.4.3). Research suggests a Likert-type scale over a continuous scale to assess the current mood and emotions of people [1]. Additionally, the complexity of input, in other words the amount and type of questions, interface and integration should be minimized as much as possible [1].

A first mock-up of the iPhone layout (see Figure 5) has been designed in the way that all the necessary information, such as a short description of the approach, an informed consent page, login or sign in possibilities and the demographic questionnaire, have been provided in the iPhone application (see Figure 5). After the login, a default screen displays saying the user needs to wait for a notification (see Appendix A). After partly implementing this first mock-up, we realized that the participant should first be informed and then be able to download the application. Otherwise the user journey itself would have had many more paths than contemplated. Therefore, the initial mock-up and user journey were adapted.

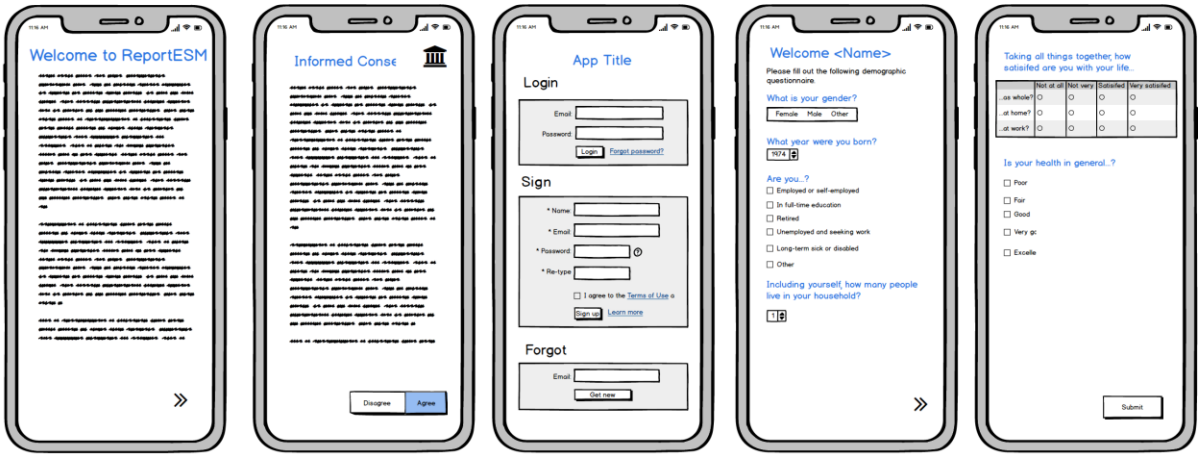


Figure 5: First Draft iPhone Design (From Left to Right: Welcome Page, Informed Consent, Login Page, First Page Demographic Questionnaire, Second Page Demographic Questionnaire)

After iterating on the initial mock-up and user journey, we focused the prototype design on the survey part. A first draft of the iPhone survey can be found in Figure 6. This draft already included a first version of the content and an implementation approach with checkboxes to minimize the complexity of the survey. The feeling questions have been designed in a table structure in order to make the selected answer more visible to the user.

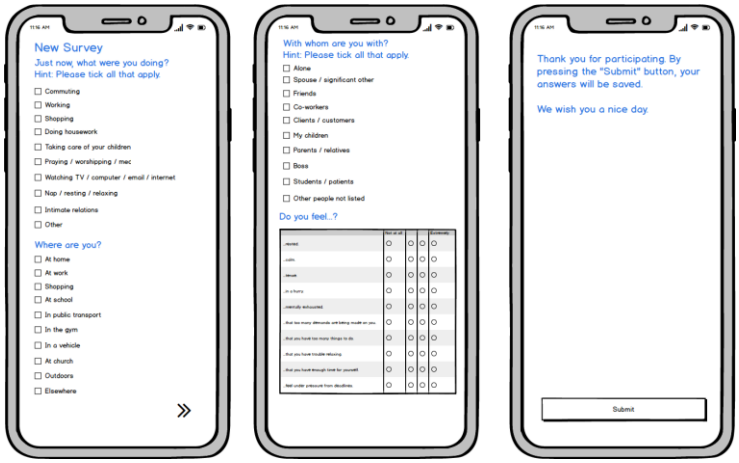


Figure 6: First Draft iPhone Survey

Due to implementation restrictions of Apple, the checkboxes in this first version could not be implemented and an approach to save the data from the table question has not been found. Iterating on this first version, resulted in many other versions including iOS checkboxes and some self-implemented checkboxes. However, these approaches did not pass the usability tests performed by the experimenter due to the fact that the user would need guidance on how to fill in the survey. Hence, this would add burden for the user to use the application. The final version of the iPhone design can be found in Figure 7 (see also Appendix A).

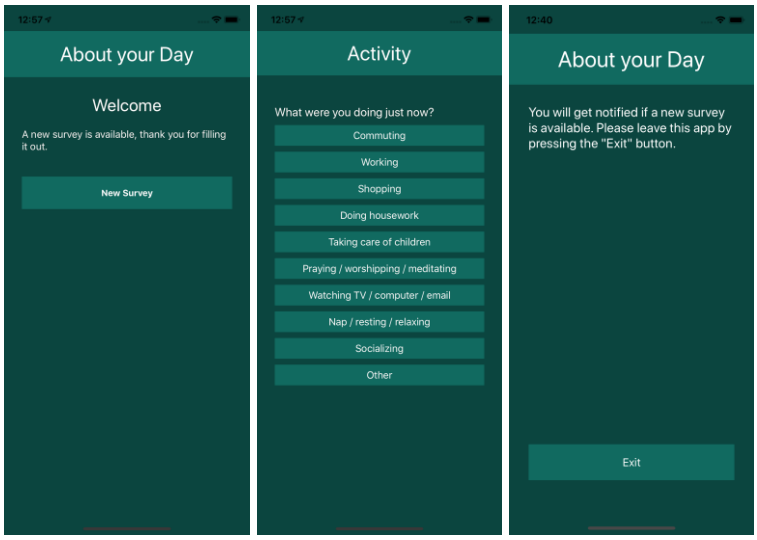


Figure 7: User Journey on the iPhone (Left: Welcome Interface, Middle: Activity Question, Right: Final Screen)

In the final version the checkboxes have been replaced with regular buttons and each question is displayed on a new interface. Additionally, research has shown that the colour green is said to be calming, natural and optimistic[1], therefore the design of the iPhone survey is green with some different nuances with white font. The colour pallet is constantly used in the whole user journey on the iPhone.

Considering the small screen size of the Apple Watch, the design of the iPhone survey could not be reused for the Apple Watch. A first design approach of the questionnaire on the Apple Watch is displayed in Figure 8. The first approach included different types of checkboxes in order to find out which one can be implemented and is most efficient for the user.

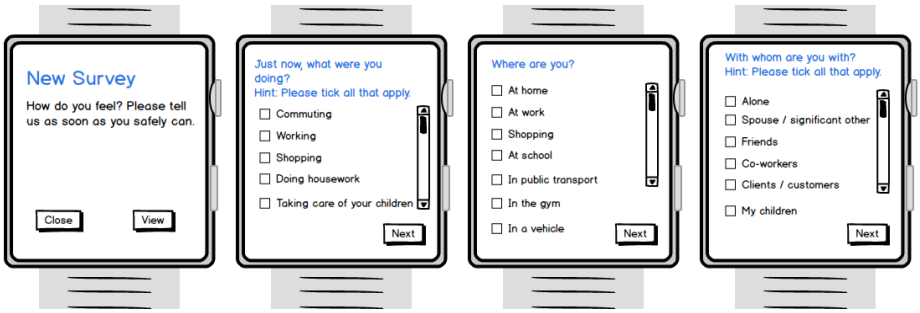


Figure 8: First Draft Apple Watch Survey

Considering the small screen of the Apple Watch various question styles have been tested to achieve the best design and usability (see Appendix A). The final version of the Apple Watch survey (see Figure 9) partly re-uses the colour pallet introduced by the iPhone survey. However, the main difference to the iPhone survey is that some questions are shown on multiple lines and long questions with a lot of selection possibilities are being displayed with a scrollable screen. Additionally, various tests conducted by the experimenter have shown that the possibility of selecting an answer by mistake while scrolling through the answers options is very high. Therefore, the Apple Watch survey has a back button, such that the user can go back and overwrite his previous answer.

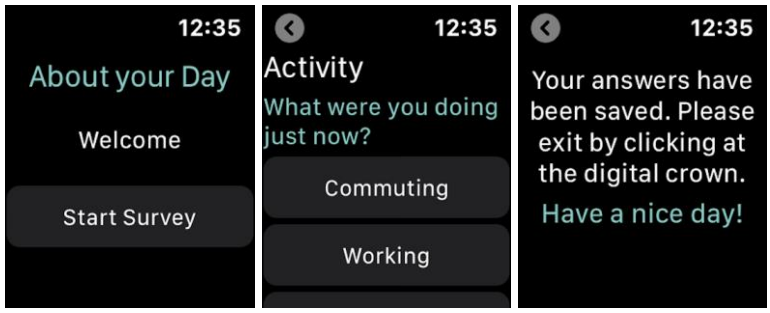


Figure 9: User Journey on the Apple Watch (Left: Welcome Interface, Middle: Activity Question, Right: Final Screen)

[1] Cherry, K. „The Color Psychology of Green" https://www.verywellmind.com/color-psychology-green-2795817

In order to improve the participants' experience with the application throughout the whole study, a logo (see Figure 10) has been created. This helps the participants to recognize the application and the information belonging to the study. The logo reflects the chosen colour pallet. However, all other elements in the logo have no significance.

Figure 10: About your Day Logo

In the end, the design of the application reflects the goal to minimize the complexity of input and provides the user with a calming and self-explanatory interface. All other screens and iterations of the iPhone and Apple Watch application can be found in Appendix A.

3.5. Questionnaire

As previously mentioned, the content of the iOS application includes a questionnaire about activities, places, the human environment and feelings. This section gives an overview over the questions, which have been used in the iOS application, their origin and the difference between the questionnaires on the different devices. Figure 11 displays the questions used in the iPhone and Apple Watch questionnaire and their corresponding answer options.

What are you doing just now?

- Commuting
- Working
- Shopping
- Doing Housework
- Taking care of your children
- Praying / worshipping / meditating
- Watching TV / computer / email / internet
- Nap / resting / relaxing
- Socializing
- Other

Where are you?

- At home
- At work
- Shopping
- At school
- In public transport
- In the gym
- In a private vehicle
- At a religious facility
- Outdoors
- Elsewhere

Who are you with?

- Alone
- Family / relatives
- Professional / work colleagues
- Friends / social acquittances

Do you feel…

	Not at all 1	Moderately 2	Very 3	Extremely 4
…rested?				
…calm?				
…tense?				
…in a hurry?				
…mentally exhausted?				
…overwhelmed?				
…overcommitted?				
…anxious?				
…balanced?				
…stressed?				
…under time pressure?				

Figure 11: Smartphone and Smartwatch Questionnaire

The first three questions about the activity, place and human environment are taken from the original ESM questionnaire [16]. The questions about the feelings have their origin in the Perceived Stress Questionnaire (PSQ) by Levenstein et al. [41]. The original survey includes 30 questions about current moods and feelings for which the participant had to mark how often this feeling applied to him during the last year or two. It further sorts its 30 items into four categories: worries, tension, joy and demands. The PSQ index is derived from raw scores, varying from 0 (lowest possible level of stress) to 1 (highest possible level of stress) [41]. It is also similar to the 14-item Perceived Stress Scale (PSS) of Cohen et al. [11], which measures the degree to which situations in one's life are appraised as stressful. The PSS is designed to tap the degree to which respondents found their lives unpredictable, controllable and overloading, which is a central component of the experience of stress. Comparing the PSQ and the PSS has shown that the PSQ leads to more subjective results than the PSS [11, 41].

Reviewing the provided coding of questions by Fliege et al. [24] and comparing the PSQ to the PSS, we decided to use the PSQ questions from the categories *tension* and *demand* to assess the level of stress people are perceiving. Their definition for the category *tension* is described by the words: exhaustion, imbalance and a lack of physical relaxation. Their definition for the category *demand* is covering: lack of time, deadline pressure and high task load [24]. Table 2

displays the reused PSQ questions, their corresponding PSQ number, the PSQ category it belongs to and the feeling used for the smartphone and smartwatch questionnaire.

No.	PSQ Question	Category	Questionnaire
01	You feel rested.	Tension	Rested
02	You feel that too many demands are being made on you.	Demand	Overwhelmed
04	You have too many things to do.	Demand	Overcommitted
10	You feel calm.	Tension	Calm
14	You feel tense.	Tension	Tense
16	You feel you're in a hurry.	Demand	In a hurry
26	You feel mentally exhausted.	Tension	Mentally exhausted
27	You have trouble relaxing.	Tension	Anxious
29	You have enough time for yourself.	Demand	Balanced
30	You feel under pressure from deadlines.	Demand	Under time pressure

Table 2: Emotions Reused from the Perceived Stress Questionnaire [41]

Additional to the emotions in column 4 in Table 2, the participants were asked if they feel stressed. More information about the stress index applied in this thesis can be found in chapter 4.5.

Comparing the questionnaire in the smart devices with the computer questionnaire has shown that there exist three small content differences. The first difference is that the feelings are assessed by using a 4-point Likert scale in the smart devices, whereas the computer survey uses a 6-point Likert scale (see Appendix B). The second difference is that in the smart devices survey the feeling questions have been shortened to "Do you feel…?" due to the small screen size. The third difference is that in the computer questionnaire the participant is more guided through inserting his reconstruction with a description at the beginning. Moreover, the user needs to state which day of the week it is, when he woke up and when he went to bed the day before. This helps to assess whether the amount of sleep influenced the responses during the assessment with the computer questionnaire. The questions and description used in the computer survey can be found in Appendix B.

3.6. Implementation Difficulties

This section shortly summarizes the major implementation difficulties, which were faced during the process and had an impact on the final product. A major impact on the notification component arose due to the fact that the development license for Apple provided by the university is restricted in terms of available certificates. The available certificates do not give any access to implement notifications in an application. However, this problem was solved by buying a private Apple developer license and gaining access to all possible certificates. Nevertheless, in order to stay on schedule with the application development a stand-alone push

notification approach could not be implemented. Another issue was faced regarding the provision of the application to the participants. Despite intensive online research with regards to the app reviewing process of Apple, no information stated that the app needs to go through a 1-2 weeks review by Apple in order to provide an app to test users. Hence, the application was submitted for this review, but was declined due to the defined narrow target group of 12 study participants. In the end we found a workaround and the participants needed another login to be able to download the application for the study.

In general, the functionalities implemented for the iPhone application could not be reused easily for the Apple Watch application, which made the implementation harder than expected. Additionally, some functional requirements had to be removed from the application, such as the login and registration page, because the logic of these requirements and the chosen basis did not match. Finally, the major difficulty encountered in developing the application lied in the small Apple Watch screen, which restricted the design possibilities for the survey in various ways, such as the type of button and the number of buttons.

4. Study Design

With the aim to answer the research questions (see Chapter 1.1), the About your Day study has been designed. The study is planned for nine days and at least six people considering counterbalancing the conditions (see Appendix B). We implemented an application with an ESM survey for the smartphone and smartwatch condition. The ESM was introduced in section 2.4.1. Details about our ESM survey are covered in chapter 3, particularly chapter 3.3 for the user journey of the application and chapter 3.5 for the ESM questionnaire itself. The web-based survey follows a DRM approach (see Section 2.4.2), which means people are asked to fill it in on their computer and reconstruct their day accordingly. The computer questionnaire has been created with a private hosted version of LimeSurvey (see Appendix B). After submitting the study design to the human subject committee of the University of Zurich and receiving positive feedback, the study took place from the 20 to 28 of May 2019. In order to increase the compliance with the methdology, participants got a detailed briefing in the form of a consent form, a schedule previous to the study, reminders during the sampling procedure and a compensation of 200 CHF at the end of the study.

4.1. Procedure

The About your Day study was designed with a within-subjects study design, meaning all participants had tested all the conditions on their private devices. The setting of the study included that participants filled in the questionnaires wherever they were when they received a notification or the invitation for the computer survey. In order to remove any order and sequence effects, we used across subjects counterbalancing (see Appendix B) and randomly assigned participants to the order of sequences. Additionally, to rule out the effect of the frequency on the participant's perception, we decided to collect self-assessments on the Apple Watch and iPhone five times per day. Moreover, the computer survey was filled in once per day. Additional to the smartwatch survey, we sensed the heart rate of the participants to gain more knowledge about the possibilities to gather heart rate data on the Apple Watch. Table 3 shows an example schedule for a participant of the About your Day study, which is explained in more detailed in the next paragraph.

Before study	**3 days**		**3 days**		**3 days**		**After study**
	Monday – Wednesday		Thursday – Saturday		Sunday – Tuesday		
Consent Form Demographic Questionnaire	Condition 1: DRM Computer	Feedback Survey DRM Computer	Condition 2: ESM Smartphone	Feedback Survey ESM Smartphone	Condition 3: ESM Smartwatch	Feedback Survey ESM Smartwatch	Overall Feedback Survey

Table 3: Example Schedule About your Day

At the beginning of the study, participants were asked to give consent (see Appendix B) and to fill in a demographic questionnaire (see Appendix B) to control for most of the external factors. In the consent form, the participants were informed that there do not exist any short- or long-term risks, discomforts or benefits. The demographic questionnaire includes questions about gender, age, employment status, number of people living in the household, an overall assessment of the satisfaction of the participants life. After submitting the demographic questionnaire, the participant was informed that the experimenter would contact him as soon as possible.

Condition 1 includes a DRM approach (see Section 2.4.2) in which participants were asked to fill in a questionnaire once per day between 5 pm and 12 pm. The invitation to the web-based survey was sent between 5 pm and 5.30 pm, and participants were asked to fill it in on their personal computer. Additionally, the participants were informed that filling in the survey takes around 30 minutes, such that they were able to schedule the assessment if necessary.

Condition 2 includes an ESM approach (see Section 2.4.1) with the smartphone. In condition 2 participants were notified at five random times per day and were asked to report about their activities and feelings. Each phone survey took around 10 minutes to be filled in. The notification schedule (see Appendix A) includes the number of notifications for each day and at which time the push notifications were sent to the participants. Additionally, the schedule for the smart devices was adapted for every day, such that participants did not always receive notifications at the same time during the three days with the specific device. This helps to prevent that participants prepare themselves to answer the survey and give pre-defined answers.

Condition 3 includes an ESM approach (see Section 2.4.1) with a smartwatch in which participants were asked to report about their activities and emotions at five random times per day. Each survey took approximately 10 minutes to be filled in. If a survey was available for the user, participants got a notification on their smartwatch, with which they could immediately fill in the questionnaire. The notification schedule for the smartwatch, varying daily, can also be found in Appendix A.

After each three days of one survey type, participants were asked to fill in a 40 minutes feedback questionnaire about the condition they had previously used. The study closed with an overall feedback questionnaire in which participants were able to rate the different conditions according to their preferences. Additionally, those participants, which were interested, received a short fact sheet (see Appendix B) with an analysis of their emotion measures.

4.2. Instrumentalization & Measures

In order to assess how different elements affect the user's experience of the data collection, we defined the influence factors comfort, intrusiveness, perceived accuracy and the effect of the device type. Each factor has been broken down into multiple smaller components (see Chapter 1.3) to capture the extent of the factor and formulate corresponding questions, which were answered by the participants in the feedback questionnaire. Most of the questions were formulated for all three conditions: computer, smartphone and smartwatch and therefore only one formulation is included in this chapter. The full list of questions included in the feedback questionnaires can be found in Appendix B. The questions, which were not asked for all device types or were only asked in the overall feedback questionnaire, have been listed with an asterisk. The factor **comfort** was divided into the subfactors enjoyableness and ease of usability. The participants' **enjoyableness** was collected with a ranking of the surveys and their general feedback. The questions used are the ones below.

- * Comparing the three different surveys, which device type did you prefer? Please rank the three types in the box on the right-hand side. The survey type at the top-most position is the one you preferred the most.
- What is your general feedback about the computer questionnaire?

The subfactor **ease of usability** was collected with the usability, naturalness and design of the survey types. The corresponding questions are listed below.

- On a scale from 1 to 5, did you understand immediately how to enter your responses? 1= not at all, 5 = totally
- * Was it more natural to complete the survey in the moment or at the end of the day? Please explain why.
- * Do you have any feedback about the design of the questionnaires?

The **intrusiveness** of the survey type was assessed with the subfactors disruptiveness, influence of the environment on the responses of the participant, effect of the time of the assessment on the responses and the ease of answering. The questions formulated are the following ones:

- On a scale from 1 to 5, how disruptive was it for you to fill out the questionnaire at the end of the day? 1 = not at all, 5 = very disruptive
- On a scale from 1 to 5, how much do you feel your environment (e.g. the people around you) influenced the way you answered the questions?
  1 = not at all, 5 = a lot

- Overall do you think the time of the day you completed the questionnaire influenced your answers? Please specify.
- Which time of the day did you find it most intrusive?
- Do you feel that completing the questionnaire in the moment influenced your answers? Please specify why and how.
- Do you feel that completing the questionnaire at the end of the day influenced your answers? Please specify why and how.

Furthermore, the **ease of answering** was made comprehensible with the subfactors duration of questionnaire, burden, understanding of questions, easiness and efficiency. For these subfactors the corresponding questions have been formulated.

- On a scale from 1 to 5, do you think the time necessary to complete a single survey was accurate? 1 = not at all, 5 = very
- On a scale from 1 to 5, how burdensome was it to fill out the surveys? 1 = not at all, 5 = very
- On a scale from 1 to 5, do you think the questions were difficult to understand? 1 = not at all, 5 = very
- On a scale from 1 to 5, how easy was it to complete the survey? 1 = not at all, 5 = very
- On a scale from 1 to 5, do you think filling out the questionnaire 5 times per day on the iPhone was burdensome? 1 = not at all, 5 = totally

In order to assess **the perceived accuracy** of the results, participants were asked about the influence of the questions on their answers, their subjective believe of the accuracy of responses, the way of asking, whether it was a representative week for them and the reason for missing an assessment. The corresponding questions in the feedback questionnaires are listed below.

- On a scale from 1 to 5, how accurately do you think your responses reflect your feelings? 1= not at all, 5 = very
- On a scale from 1 to 5, how much did the way you were asked about your feelings (i.e. phrasing of the questions) influence the way you answered these questions? 1= not at all, 5 = very much
- Did filling out the questionnaire influence how you felt? Please specify why.

- On a scale from 1 to 5, were these representative 9 days of your life? 1= not at all, 5 = very
- If you were sometimes unable to respond, what were the main reasons?

The **effect of the device type** was assessed with an analysis of the quantitative data and the following question.

- On a scale from 1 to 5, did the form of the survey, i.e. that you answered the questions on your computer, influence how you answered the questions? 1 = not at all, 5 = totally

4.3. Recruiting

Participants were recruited with a flyer (see Appendix B), which was distributed at the University of Zurich, via e-mail and personal contact. Before the study, each participant was informed about the study purpose in general, the procedure, and about the usage of the captured data. The study purpose in general means that participants have not been informed about the fact that the study mainly focuses on the self-reporting of stress. Participants were informed that the study purpose was to find the best way to gather self-reported data about activities and emotions, because telling participants that the study focused on stress would result in biased measures. Participants had to be older than 18 years, fluent in English, available for the nine days of the study, have the possibility to respond in the moment and have an iPhone and an Apple Watch. The recruiting resulted in 12 participants and no opt-outs during the whole study. Before the study, all participants were asked to fill in the demographic questionnaire. The results are summarized in Table 4.

	Gender	Age	Education	House	Satisfaction Overall	Satisfaction personal life	Satisfaction professional life
P01	Female	31	Employed	Partner	Satisfied	Satisfied	Satisfied
P02	Male	30	Employed	Partner	Satisfied	Very satisfied	Not very satisfied
P03	Female	25	Employed + studying	Partner	Satisfied	Very satisfied	Satisfied
P04	Male	27	Employed + studying	Partner	Satisfied	Satisfied	Very satisfied
P05	Female	26	Employed + studying	Flatmate	Very satisfied	Very satisfied	Very satisfied
P06	Male	29	Employed + studying	Partner	Not very satisfied	Satisfied	Satisfied
P07	Male	24	Studying	Flatmate	Satisfied	Satisfied	Very satisfied
P08	Male	30	Studying	Flatmate	Satisfied	Satisfied	Satisfied
P09	Male	24	Employed + studying	Partner	Satisfied	Satisfied	Satisfied
P10	Male	25	Employed + studying	Alone	Very Satisfied	Very Satisfied	Very Satisfied
P11	Male	40	Employed + studying	Partner	Not very satisfied	Not very satisfied	Not very satisfied
P12	Female	29	Employed	Flatmate	Satisfied	Satisfied	Not very satisfied

Table 4: Results of the Demographic Questionnaire

Out of the 12 participants, four were female and eight were male. Besides two participants who are only studying, all participants are either employed full-time or part-time. Only one of the participants is living alone and the average age is 28 years. Overall participants are mostly satisfied with their life in general and their personal life in the past year. Four participants are very satisfied, five participants are satisfied and three are not very satisfied about their professional life.

4.4. Data Collection

The data collection was organized from 20 to 28 May without any restriction to a specific place and never getting in personal contact with the participants. The data gathered in the About your Day study consist of three different types. The first type of data is the collected feedback of the participants regarding the comfort and enjoyableness, intrusiveness, ease of answering, ease of usability, perceived accuracy of feelings and effect of the device type. An overall assessment of the survey types was surveyed on the last day of the study. As the feedback questionnaire included open-ended questions and Likert-scale questions, these measures are of quantitative

and qualitative nature. The second type of data includes the computer, smartphone and smartwatch survey as itself, which covers data about the activity, place, human environment and feelings of the participants. The collected data from these surveys represent quantitative and qualitative data. The third type of data we collected, are the heart rate measures of each participant while they were filling in the survey on their Apple Watch. This data is of quantitative nature but is not going to be analysed within the scope of this thesis.

Due to technical reasons, some participants answered the questionnaire on their Apple Watch more than requested. This additional data has not been excluded from the analysis and therefore a response rate of a participant could reach over 100%. In the end, all the collected data was saved on a server hosted by the University of Zurich (see Appendix A) and accessible by the examiner. All the data was exported into .csv-Files or Excel Sheets and processed into either R, for the daily survey on the computer, smartphone and smartwatch, or Excel, for the qualitative data of the feedback surveys.

4.5. Analysis

The quantitative analysis was performed using R for Windows version 3.5.0. The qualitative analysis has been performed using Excel and Word. The first type of data from the feedback questionnaires was analysed using the procedure described below as well as a generic grouping mechanism by marking comments, which contain the same feedback, and assigning them into the same cell in the table. The raw comments of the participants have been adjusted to remove any minor misspelling mistakes and improve the readability. The additional qualitative data of the feedback questionnaires was not cleaned or processed before the analysis. We consider that each ranking question had a scale from 1 to 5, which results in the fact that the minimum sum given for a question could be 12, when all participants responded with a 1. In contrast, a maximum sum of 60 was reached, when all participants responded the scale question with a 5. Hence, the new spectrum of sum values was again categorized into 5 groups: values of 12-21 are a rating of 1, values of 22-31 are a rating of 2, values of 32-40 are a rating of 3, values of 41-50 are a rating of 4, finally, values of 51-60 are a rating of 5. Since the spectrum of 48 values is not dividable by 5, the rating 3 is only represented by 8 numbers instead of 9 numbers, because research has shown that people tend to select more often a value in the middle instead at the edges of the Likert-scale. For each question the individual ratings were summed up and an overall rating was assigned using the categorization described above.

The factor efficiency was calculated by subtracting the start time from the end time for each assessment participants filled in and then grouped by survey type. Additionally, each assessment was linked to the specific notification the participant responded to with the aim to calculate how fast participants responded to the notification (see Appendix C).

The second type of data, the quantitative and qualitative results of the computer, smartphone and smartwatch survey, was pre-processed to answer the question about the influence of the device type and to provide the participants with a small fact sheet (see Appendix B). First, the data was made anonymous by renaming the user with P01 to P12 throughout all the datasets. Then the quantitative data has been formatted for the analysis by renaming various variables, for example the activity, place and human environment measures. The time of each response was assigned to a label to make it more feasible for the participant. All responses from 7 am to 10 am are labelled morning, those from 11 am to 1 pm are labelled noon, responses from 2 pm to 5 pm are labelled afternoon and responses from 6 pm to midnight are labelled evening. The quantitative data of the computer survey has been formatted such that the 6-point Likert scale can be compared to the 4-point Likert scale of the smartphone and smartwatch survey.

In order to measure the stress level of participants, the measures of the feelings were aggregated to obtain a measure of the subject that is collapsed across time and multiple measures. The stress rating includes the aggregated value of all eleven feelings: calm, balanced, rested, tense, overcommitted, overwhelmed, stressed, mentally exhausted, in a hurry, under time pressure and anxious. In order to add the measures of the positive feelings (calm, balanced and rested) to the aggregated value, they were converted to an inverse feelings scale to be comparable to negative feelings. Therefore, the maximum stress rating value is 44 for one participant and the minimum is 11. The higher the stress rating value, the more stressed is a person. For the hypothesis on the impact of the device type, the mean and standard deviation for each stress rating grouped by device and day was calculated and plotted. As mentioned before, the third type of data, which is the collected heart rate, was not analysed within the scope of this thesis. However, the heart rate measures have been pre-processed to be ready for an analysis. In the end, the three different datasets from the computer, smartphone and smartwatch survey were identically formatted such that they could be combined, after correcting wrong day and time label, and filtered for each participant.

5. Results

This section entails the results obtained from the About your Day study described in Chapter 4. The pre-processed results can be found in Appendix C and the raw results on the attached CD- ROM. The total number of possible responses for each participant for the whole study is 39 assessments and four feedback questionnaires. Participants responded on average in 88.7%, whereas three participants responded to 39 or more assessments. The response rate for the four feedback questionnaires resulted in 100%. During the assessments, participants were most of the time *working*, *at home* and *alone*. The activity performed the second most was *other* and the place where participants were the second most was at their *workplace*. The human environment, i.e. the person, the participants did spend time with the second most was with their *professional and work colleagues*. For more insights regarding the activity, place and human environment answers see Appendix C. In order to understand the results in this section, chapter 4.5 Analysis is prerequisite.

5.1. Comfort: Enjoyableness

The enjoyableness of participants to gather self-reported stress data has been measured with a ranking question of the different survey types and their general feedback about each survey type. The results of these measures are described in these sections.

5.1.1. Ranking of Surveys

Each participant was asked to rank the three different survey types according to their perceived comfort and preference. Table 5 contains the ranking for each participant on the first place and the sum of participants which have selected the specific questionnaire as in their opinion the best. Table 6 displays the results for each participant for the surveys on the second rank and the number of participants which have selected the specific survey as the second best. Table 7 includes the number of participants which have ranked the specific survey on the third rank and their individual rating. Participants which have selected the same combination for their ranking have been coloured in the same way. Participants giving the smartwatch questionnaire the first rank, the smartphone questionnaire the second rank and the computer questionnaire the third rank have been coloured green. Participants selecting the computer questionnaire on the first rank, the smartwatch questionnaire on the second and the smartphone questionnaire the third rank have been coloured dark green. The participants with the light green colour are those which have ranked the smartphone questionnaire the first, the smartwatch questionnaire second and

the computer questionnaire third. Those participants which have not been coloured, could not be grouped with another participant.

Survey Type						Sum
Computer Questionnaire	P1	P3	P4	P8		**4**
Smartphone Questionnaire	P7	P11	P12			**3**
Smartwatch Questionnaire	P2	P5	P6	P9	P10	**5**

Table 5: Results Rank 1

Survey Type							Sum
Computer Questionnaire	P7						**1**
Smartphone Questionnaire	P2	P5	P6	P8	P9	P10	**6**
Smartwatch Questionnaire	P1	P3	P4	P11	P12		**5**

Table 6: Results Rank 2

Survey Type								Sum
Computer Questionnaire	P2	P5	P6	P9	P10	P11	P12	**7**
Smartphone Questionnaire	P1	P3	P4					**3**
Smartwatch Questionnaire	P8	P7						**2**

Table 7: Results Rank 3

The corresponding reasoning to the participant's ranking can be found below. The comments have again been grouped according to their ranking and the title corresponds to the ranking in the Tables 5-7.

Feedback for the ranking combination: Smartwatch, Smartphone and Computer (green)

- **P2:** "iWatch was the best since it is very subtle, and I had finally something to really do with the watch where it is usually just showing information where the full story has to be read on the phone. The phone was also ok. but less handy than the iWatch, I also preferred that there was a back button in the watch app which did not exist in the iPhone app. I really didn't like the Computer app since there everything seemed so distant and for me problematic was that the previously answered questions were not visible but for me it would have been much easier since usually I remembered when I had a good time and a bad time like a mood curve but without knowing what I previously answered it seemed a bit random for me."
- **P5:** "I did not really like the computer survey and would not use it in my routine. I could imagine using the watch survey. My main problem with the computer survey was, that it is a reflection over the entire day and not on the current situation. I liked the watch survey better than the phone survey because I immediately felt the notification and could answer it without searching the phone in the bag and could easily and fast answer the questions. I was happy about the watch notifications when using the phone survey, but if this would

not have been notified on the watch as well, I might not see the notification on the phone. Because I usually have my phone in night shift mode which mutes all notifications and turns off vibrations."

- **P6:** "Convenience. You are not using your phone for everything if you have a watch. For the iPhone survey you sometimes have to take your phone into your hands just to complete the survey - which is burdensome. The Computer survey in the end of the day did not feel well, at least for me."
- **P9:** "The watch was sooo much faster for me. I did not need to pull out my phone or really interrupt what I was doing. The phone was similar but not quite as convenient, as I do not always have my phone with me during the day. The computer survey was the most annoying for me and felt like it takes the longest (even if that probably is not true)."
- **P10:** "Filling out the survey on the watch is less interruptive than on the phone. Especially, when I was with a group of people. Filling it out on the watch and explaining what it was, was interesting to my colleagues. However, getting the phone out seemed more like I don't want to interact with them, and it was considered ruder."

Feedback for the ranking combination: Computer, Smartwatch and Smartphone (dark green)

- **P1:** "I could take my time to fill out the computer survey in the evening and think my answers were more accurate then. Since I always have my watch with me, I could do the Apple Watch survey immediately and it was easy to handle I missed some notifications on the phone easily during the working days."
- **P3:** "The computer survey was easily included in my schedule as it was once per day. I like the iPhone survey the least because when the pop up appeared on my watch, I had to immediately look for the phone. The problem was solved by the watch survey which I put as a second choice."
- **P4:** "I like to think about my answer (and change possible inputs). But I guess the questionnaire is intended to capture emotions that are more automatic and not deliberative."

Feedback for the ranking combination: Smartphone, Smartwatch and Computer (light green)

- **P11: "**iPhone: good in accuracy and receiving notification. Apple Watch: good in accuracy, bad in receiving notification. Computer survey: bad in accuracy (recalling emotion is not easy)."

- **P12: “**All the answer possibilities could be better seen on the phone. the display of the watch is a bit too small for that, however, on the watch through the vibration I realized it more soon when e new survey was available.”

Feedback for the ranking combination: Computer, Smartphone, Smartwatch

- **P8:** “From a UI/UX perspective I found the iPhone survey the most comfortable one. It was very quick and easy to fill in the survey. However, I preferred the computer survey a bit more simply because it was just one survey a day and it was in the evening when I usually had time to fill it in. It felt less interrupting than the iPhone surveys.“

Feedback for the ranking combination: Smartphone, Computer, Smartwatch

- **P7: “**iPhone was the most natural way to answer as it was quickly available and convenient to answer. The computer survey was convenient as well but I think my answers don't truly reflect my feelings at a certain point since it was hard to think back how I felt at a certain point. The apple watch survey was slightly annoying as I had to keep it on my wrist all day long and answering questions on such a small screen was sometimes annoying (especially at the gym or in public transport).”

The rankings have been summarized in Table 8 and the survey type which got the most ratings for the specific rankings has been marked grey. We can see that the smartwatch questionnaire has been ranked the most on the first rank, the smartphone questionnaire the most as the second rank and the computer questionnaire the most on the third place of the rank. In other words, participants liked the smartwatch questionnaire the most, the smartphone questionnaire the second most and the computer questionnaire the least.

Survey Type	Ranking 1	Ranking 2	Ranking 3
Computer Questionnaire	4	1	**7**
Smartphone Questionnaire	3	**6**	3
Smartwatch Questionnaire	**5**	5	2

Table 8: Results of the Ranking

5.1.2. General Feedback Computer Questionnaire

The participants have also been asked to give general feedback on the computer questionnaire. The results of this optional question are presented in this section. As this question was optional, not every participant has provided his or her opinion about the computer questionnaire.

- **P1:** “It is easy to understand the way it works and didn't take a lot of time. I found it difficult to scale my feelings and I'm not sure how accurate it would be.”

- **P3:** “The computer survey was my favourite one as I could easily include it into my daily schedule. However, in comparison to the watch and iPhone surveys, the results could not portray my emotions so specifically as I sometimes had difficulty recalling feeling particular emotion or state.”
- **P4:** “It was not possible to specify different types of people I met during the 3 hours (3 hours are quite a long time).”
- **P5:** “I would not use the computersite to answer questions about my day. Because when I am away from work, I usually do not use a laptop/computer and mainly that's why I would not use it. Moreover, I did not really like to reflect every day at the end of the day over the entire day and put each situation and feeling in perspective (sich nochmals in die Situation versetzen).”
- **P6:** “As said before: Answering these questions in retrospective is an estimate and mobile/app are probably more accurate. Usability of the computer survey is good, but having to fill it out every evening - where also other things have to be done - is a little tiresome.”
- **P7:** “It was interesting, but I regret that the time scope did not go through the evening as I am convinced that feelings in the evening as really important as well. Maybe include "studying" when asked "what were you doing at that time" there is mention for that.”
- **P8:** “I found this one probably the least intrusive one. First, because it was only one survey a day and second, because I usually had time in the evening to fill in the survey and didn't feel interrupted by it as much as the Watch/iPhone surveys.”
- **P9:** “It just seemed too take longer, even-though objectively it probably did not take as much time as the phone surveys of a day combined.”
- **P10:** “I find the 6-point Likert scale (if I remember correctly) difficult to answer. For me, I am either stressed, not stressed or in the transition. This transition, however, is very difficult to judge for me. I wish it was more like a 3-point or 4-point scale rather than 6-point.”
- **P11:** “Recalling my feeling during the day was way more difficult than I thought.”
- **P12:** “The questionnaire is self-explanatory, easy to understand and easy to fill in.”

The results above have been clustered into positive and negative feedback to get a better overview of what participants liked and what they did not like about the computer survey. Any identified wishes or suggestions in the feedback will be resumed in chapter 6. Table 9 displays the clustered feedback, where similar statements have been placed in the same table cell.

What is your general feedback about the computer questionnaire?	
Positive Feedback	**Negative Feedback**
• It was interesting	• Feelings in the evening are important too
• Easy to understand the way it works • Easy to fill in • Good usability • Self-explanatory	• 6 - point Likert scale difficult to answer • Transition of feelings is difficult to assess • Difficult scale feelings in general
• Does not take a lot of time • Seemed to take longer compared to the phone, but it does not	• Difficult to recall feeling a particular emotion or state; accuracy? • Answering the questions in retrospective is an estimate
• Easy to understand questions	• Filling it out in the evening is tiresome
• Least intrusive: once per day, more natural to include in schedule & did not feel interrupted • Easy to include in schedule	• Would not use computersite to answer questions about my day → only use computer at work • Hassle of getting laptop

Table 9: Results of the General Feedback about the Computer Questionnaire

Overall participants were satisfied with the computer questionnaire and especially its usability. It was easy for them to understand how it works and some participants mentioned that it does not take much more time than the smartphone and smartwatch questionnaire. However, it was very hard for the participants to assess their feelings on a 6 – point Likert scale, answering the questions in a retrospective is just an estimate and for some participants their computer is negatively associated to work.

5.1.3. General Feedback Smartphone Questionnaire

Additional to the ranking question participants have also been asked to state their general feedback on the smartphone questionnaire. The results of this optional question are presented in this section. As this question has been optional, not every participant has provided his or her opinion about the smartphone questionnaire.

- **P1:** "It was easy to handle. Nevertheless, I think 5 times a day are too much."
- **P2:** "The iPhone part should have a back button. When the app is closed during the survey it can't be finished."
- **P3:** "I believe the questionnaires had a positive effect on me. After answering some of the questions, I took my time to think about particular emotions being connected to the place I am right now, the people (or lack of them) around me as well as the activity I was performing. Filling the survey on my iPhone was always quick. Notifications were a nice

way of reminding me about the survey as keeping track with a paper schedule would be problematic."

- **P4:** "I think the questionnaire was easy to understand and fill in. However, I missed the back button. Moreover, I encountered 2 unintended exits from the app (Absturz)."
- **P5:** "Unfortunately, I had the situation that I was interrupted during the survey by another person, and then my phone went into the lock mode and when I wanted to continue the survey, it was already submitted and I could not complete it entirely. Before I noticed this, it already happened a few times, I am sorry for that. Overall, I did not really like to fill out the questionnaire on the phone. I guess it is mainly because I do not really want to be on the phone during work. Moreover, I usually have the phone on the 'night mode' which turns off all vibrations, in order to not disturb my colleagues. Therefore, I really liked that the notifications also popped up on the watch and I saw that I had to fill out a new questionnaire."
- **P6:** "The app was easy to use, no errors or whatsoever. Certainly, a good alternative to the watch version."
- **P7:** "I think you should have asked questions about how I felt during the evening as well. Otherwise, it was really easy to follow the questions and respond in time."
- **P8:** "I found the questionnaires on the iPhone more user-friendly than on the Watch, simply because the interface was easier to use with the bigger screen and the better responsiveness of the app. I did miss more questions in the iPhone session because I was out with friends a lot during this time and didn't see the notifications or didn't want to interrupt the interactions with my friends."
- **P9:** "Doesn't always vibrate when a new survey is there, so I missed it sometimes."
- **P10:** "The iPhone questionnaire was much nicer compared to the computer version. Filling it out on the phone is a lot quicker. However, when I was in a conversation it seemed to be rude to answer the questionnaire on the phone. With the watch this seemed to be less of an issue."
- **P11:** "It was okay in general. Maybe it's because of the app, but sometimes I received multiple notification together and what I did was to answer them one after the other in ten minutes or so. I am not so sure if that's the expected behaviour."
- **P12:** "It is really easy to fill in and only takes a few seconds. If you press the wrong button, you cannot go back, that is sometimes a challenge, but it is better as you cannot take too much time to correct your answers."

The results above have again been clustered into positive and negative feedback to get a better overview of what participants liked and what they did not like about the smartphone survey. Any identified wishes or suggestions in the feedback will be brought up in the chapter 6. Table 10 displays the clustered feedback, where similar statements have been placed in the same table cell.

What is your general feedback about the phone questionnaire?	
Positive Feedback	**Negative Feedback**
• Easy to fill in • Easy to use, no errors or whatsoever • Easy to understand • Easy to follow the questions	• When app is closed during survey, the survey cannot be finished • If you press the wrong button you cannot go back
• Survey had positive effect, started to reflect about my emotions being connected to the current place, people around me and my current activity	• Seems more rude to answer survey on phone than on Watch in presence of others • Do not want to be on Phone during work
• More user-friendly than on the Watch: easier to use interface & better responsiveness of app • Good alternative to the watch	• Does not always vibrate when new survey is available • Application did crash • Received multiple notifications together
• Much nicer compared to the computer version • Lot quicker than the computer • Notifications are a nice way to remind me	• 5 times a day too much
• Notifications do also pop-up on Apple Watch	
• Easy to respond in time • Easy to handle • Very fast	

Table 10: Results of the General Feedback about the Smartphone Questionnaire

The feedback shows that the participants were very satisfied with the smartphone questionnaire. They see the smartphone questionnaire as a good alternative to the watch questionnaire and it was easy for them to interact with the smartphone questionnaire. The negative feedback is mainly regarding technical functionalities, such as that if the app is closed during the self-report, the survey cannot be finished anymore and regarding the negative connotation of the smartphone in presence of others.

5.1.4. General Feedback Smartwatch Questionnaire

Participants were asked to write-down their general feedback on the smartwatch questionnaire to complement the ranking question. The results of this optional question are described in this section. As this question has been optional to answer, not every participant has provided his or her opinion about the smartwatch questionnaire.

- **P1:** “Easy to handle, less disruptive than other questionnaires.”
- **P2:** “The scale from not at all to extremely was a bit strange, positive feelings sound strange together with the word extremely ‘extremely calm’ sounds somewhat negative for me, like a little too calm.”
- **P3:** “I preferred the Apple Watch surveys over the iPhone ones as they were easier to fill. Overall, I did not notice a change in my emotional state influenced by a particular device.”
- **P4:** “It was better as expected. Sometimes it did not send my inputs, or I was not sure whether it did. Despite that it was very easy to fill in and captured my emotions of the moment.”
- **P5:** “I was not sure if this functionality exists, but I had the impression not: If one has no internet connection (or another failure to send the survey), the survey should be stored locally and send as soon as the internet connection is available. Otherwise, I have to remember to open the App again and submit the survey.”
- **P6:** “Interesting, but a little too often in my sense. The menu was easy to use (on a 44mm apple watch), so: Well done! Sometimes I did not get notifications but realized a while later that there has been a survey available, just fyi.”
- **P7:** “This was the less effective questionnaire of the three (computer, iPhone, watch) because I didn't necessarily had my watch on my wrist all day long and sometimes the notification appeared on my phone but not on the watch so I had manually open it in the watch. Also filling a questionnaire on such a small screen is not really convenient. However, I think this could be useful to use this apple watch to measure heartbeat rate at the time of the questionnaire (which we can't do on the iPhone).”
- **P8:** “I think it worked quite well, since the interaction with the watch is simple and short. I would probably still prefer filling in the survey on the iPhone because the UI is bigger, and my iPhone is faster.”
- **P9:** “I think once the technical issues (vibration and that it pops up while the hr is being recorded) are solved, it will be the best version to use. At least for me, as I don't like having my phone around the whole time, so being able to do everything on the watch is great.”

- **P10:** “Generally speaking, it was very lightweight and easy to do. However, I made some mistakes. For example, "feeling anxious" --> "not at all" which is on the left. "calm" --> "extremely" on the right. I sometimes accidentally pressed the button on the wrong side. There was no possibility to correct this error.”
- **P11:** “It was a bit tedious because I didn’t always receive the notification via the watch. Often I see several notifications on my iPhone once together.”
- **P12:** “It was really easy to answer, short questions, only a few questions. Sometimes the system was not answering immediately so I pressed submitted more than once. In general, it was not time consuming.”

The results above have again been clustered into positive and negative feedback to get a better overview of what participants liked and what they did not like about the smartwatch survey. Any identified wish or suggestion in the feedback will be discussed in the chapter 6. Table 11 displays the clustered feedback, where similar statements have been placed in the same table cell.

What is your general feedback about the watch questionnaire?	
Positive Feedback	**Negative Feedback**
• Menu was easy to use • Worked quite well • Lightweight & easy to do • Easy to handle	• Prefer iPhone: faster and bigger UI • Filling out questionnaire on small screen is not convenient
• Better as expected • Interesting • Do everything on watch is great	• The system did not always answer • Sometimes submissions were not send / was not sure if they were sent
• Simple and short interaction • Took only a few seconds • Short questions • Easy to answer	• Least effective one compared to computer and phone (did not always wear Watch and had to manually open app on Watch) • Did not always get notifications
• Prefer Watch surveys over iPhone: easier to fill in	• Too often
• Useful to measure heartbeat rate during the survey • Captured emotions of the moment	• Was confusing to have scale switches for the feelings balanced, relaxed and calm → wrong answer has been selected
• Less disruptive than others	

Table 11: Results of the General Feedback about the Smartwatch Questionnaire

The participants were overall very satisfied with the smartwatch questionnaire, because it was lightweight, easy to handle and a simple, short interaction. The negative feedback was mainly

regarding the functionalities on the smartwatch, such as its reliability, and regarding the small user interface.

5.2. Comfort: Ease of Usability

In order to assess the ease of usability of the application and different survey types, the factor was broken down into questions regarding the usability, input type, naturalness and design of the application and questionnaires (see Chapter 4.2).

5.2.1. Usability

On a scale from 1 to 5, did you understand immediately how to enter your responses? 1= not at all, 5 = totally		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	56	5
Smartphone Questionnaire	56	5
Smartwatch Questionnaire	56	5
Overall Questionnaire Rating	58	5

Table 12: Results Usability

In all questionnaires the participants rated the question that they immediately understood how to enter their responses with a 5 (totally). The additional comments have shown that participants had no problems with the surveys, that the questionnaires were very clear and easy to use.

5.2.2. Naturalness

Participants were asked whether to respond in the moment was more natural than to respond at the end of the day. The results of this question are listed below.

- **P1:** “It was more natural completing the survey at the end of the day. I could take my time to focus on the survey and reflect my thoughts.”
- **P2:** “It is much better to do it in the moment, since one has to think about the day and decide how one felt at a specific point in time, but for me most of the time this was very difficult or I just remembered a vague mood curve.”
- **P3:** “In the moment. Recalling particular emotions from the morning hours while filling the survey in the evening was sometimes challenging.”
- **P4:** “I would not say natural, but perhaps more spontaneous and more honest - however I did sometimes mistakes on the phone or watch and could not go back.”

- **P5:** “Definitely in the moment. If I had to think about the day and rate the feelings from the morning, I felt that my answer might have been influenced by the feelings which happened afterwards (e.g. the afternoon). Moreover, for me it was hard to look at the feelings and rate them depending on their time-occurrence. e.g. just think about the feelings in the morning. Except of the feeling which were very intense and remarkable, but not a usual feeling.”
- **P6:** “As explained, I liked the ones in the moment better.”
- **P7:** “In the moment because it was way easier to truly report my feelings.”
- **P8:** “I would say recording my mood in the moment several times throughout the day gives me a better insight into my emotions, because emotions can change quite drastically in the course of a day. Also, when filling in the survey in the evening/end of the day I might not remember exactly how I felt in the morning. Also, the questions in the end-of-day survey spanned over three hours, which made giving answers more difficult because I might have felt very calm at the beginning of those three hours but felt stressed at the end of them.”
- **P9:** “Definitely in the moment! I was able to just put in how I was currently feeling and did not have to think back and try to remember how I was feeling.”
- **P10:”** I prefer giving it right at the moment. This way the memory is still fresh, and most of the times it was not really interruptive. However, when I was sitting in a meeting, I filled it out right after.”
- **P11:** “The moment, because of the accuracy of emotion.”
- **P12:** “As it pops up on the screens like massages etc. I think we are conditioned to answer it immediately. So, I found it natural to answer it immediately, like a task, hence, I had one thing done from my „to do list“.”

In each of the comments above it has been identified whether participants preferred to respond in the moment or at the end of the day. The grouped results of the identification are displayed in Table 13.

Was it more natural to complete the survey in the moment or at the end of the day? Please explain why.	
	Number of participants
Right in the moment	11
Reconstruct the day	1

Table 13: Results Naturalness

Besides one participant, all participants stated that it was more natural to complete the survey in the moment than at the end of the day. The most mentioned reason is that participants felt it was challenging to fill out the survey at the end of the day and responding right in the moment made them respond more honest. One participant added that reporting his current several times throughout the day gives him a better insight into his emotions, because in his opinion emotions can change quite drastically over the course of the day. Another participant described that he felt that his answers in the computer questionnaire for the morning slot might have been influenced by the feelings which happened in the afternoon.

5.2.3. Design

Each participant was asked what he or she likes or dislikes about the design of the questionnaire. The results of this question are listed below.

- **P1:** “Some adjectives are pretty the same (e.g. calm and rested) and it was difficult to differentiate.”
- **P2:** “My feeling is that most of the feelings asked are more negative than positive. The four choice questionnaires were much easier to answer then the one with 7 possibilities. The Computer questionnaire was most difficult and annoying to fill in if all these questionnaires are compared. I preferred the iWatch since it is the most subtle way to fill in the questionnaire and it is very handy since no device has to be picked out of the pocket it can just be done, when ever needed.”
- **P3:** “It was simple and straightforward which made filling the survey quick. Various pop ups served as a great reminder. The only thing changing were the adjectives which saved time on reading the whole sentences all over.”
- **P4:** “Use the same scale across input types extremely calm sounds awkward to me.”
- **P5:** “Yes. In the question 'Were you with anyone' I missed the answer option 'with random people', because sometimes I was surrounded or had to deal with persons which were neither my colleagues, friends or family.”
- **P6:** “Sometimes it was quite hard to assess the right answer as it's quite a jump between "moderate" and "not at all" (or the other). I'm quite sure that you covered the why in the design of your questionnaire, but I wonder if an additional answer (or two) wouldn't help.”
- **P7:** “In general, it was really good, but I was expecting a more in-depth feelings analysis (more questions about other sentiment traits).”

- **P8:** “It so happens that I was eating a few times when filling in the survey, but eating wasn't one of the answers in the activities question. I selected "other" instead.”
- **P9:** “I think it was a good amount of questions. However, I did think that filling out the same questions for different times of the day on the desktop survey was a bit boring.”
- **P10:** “I gave some feedback above.”
- **P11:** “The scale is a bit vague.”
- **P12:** “It would be helpful to go back if one pressed something wrong, but on the other side it is good that one cannot go back to complete the survey in a reasonable time. Otherwise I find the design simple to understand and easy to handle and clear.”

Each of the comments above has been analysed and broke down into single pieces of feedback participants have been given. Each piece has then been identified whether it is positive, negative, a wish or suggestion. Once again, the wish or suggestion will be discussed in the chapter 6. The clustered findings can be found in Table 14.

Do you have any feedback about the design of the questionnaire?	
Positive Feedback	**Negative Feedback**
• Simple and straightforward • In general, really good • Simple to understand • Clear • Easy to handle	• Hard to assess the right answer: jump between “moderate” and “not at all” • Use the same scale across input types • Scale is a bit vague • Expected an in-depth feelings analysis
• Only changing the adjective: saved time • Very fast to fill out survey • Good amount of questions • Pop ups served as great reminder	• Feelings were similar; difficult to differentiate • More negative than positive feelings have been asked
• 4-point scale much easier to answer than 7-point scale	• Computer questionnaire was most difficult and annoying • Filling out same questions for time intervals on the computer questionnaire was bit boring

Table 14: Results about the Design

The feedback about the design of the questionnaire has shown that it was perceived as simple, straightforward and the pop ups served as a great reminder. The participants criticized the content of the questionnaire, such as the labelling of the scale and the selection of feelings.

5.3. Intrusiveness

The intrusiveness of the survey types has been assessed with the factor disruptiveness, influence of the environment and timing of the self-reports (see Chapter 4.2).

5.3.1. Disruptiveness

On a scale from 1 to 5, how disruptive was it for you to fill out the questionnaire at the end of the day/during the day? 1= not at all, 5 = very		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	23	2
Smartphone Questionnaire	37	3
Smartwatch Questionnaire	29	2

Table 15: Results Disruptiveness

The smartphone questionnaire has been rated as the more disruptive questionnaire compared to the computer and smartwatch questionnaire. Its disruptiveness is moderately, whereas the disruptiveness of the computer and smartwatch questionnaire is small. Some participants mentioned in specific that the computer questionnaire was more convenient in the aspect of filling in one survey per day compared to a few during the day. Additionally, it was also more convenient to fill in the questionnaire in the evening, when the participants had time to fill in the survey. Others commented that getting their laptop was a hassle and this led them back to the feeling of working.

Out of the nine comments regarding the smartphone questionnaire, five comments mentioned that the notifications sometimes occurred at times when they were not able to answer. For example, in meetings or during a conversation with friends. From their point of view, the survey as itself was not much of an issue, but rather the fact that they had to stop what they were doing and then to restart again. Two participants stated specifically that they felt the smartphone questionnaire was more disruptive than the smartwatch, but less than the computer version. Four participants commented on the smartwatch questionnaire that they had the feeling that the computer survey did interrupt their workflow and postponing the notification on the smartwatch was hard, because they mostly forgot about it and the notifications disappears after some time. On the other hand, four participants described that since the survey was very quick on the smartwatch it did not really bother them to fill it out.

5.3.2. Environment Influence

On a scale from 1 to 5, how much do you feel your environment (e.g. the people around you) influenced the way you answered the questions? 1= not at all, 5 = a lot		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	27	2
Smartphone Questionnaire	25	2
Smartwatch Questionnaire	27	2

Table 16: Results Environment Influence

All three surveys have been rated at the same level regarding their environmental influence. The environmental influence is small for all the three types. Four participants specified that they were mostly alone when they were answering the questions and therefore the environment had a small or no influence on the way they answered. Two other participants stated that they would not know how the environment could affect their feelings and therefore rated the environmental influence on answering as not at all. In the end, various participants noted that they tend to answer more positively in presence of others and that it definitely had an influence on their emotions, but not on the way they filled out the survey.

5.3.3. Timing of Self-reports

Participants were asked in the overall questionnaire whether they think that the time of the day had an influence on how the responded in the self-reports (see Appendix B). The results of the open-ended question are described below.

- **P1:** “I don't think so.”
- **P2:** “Whether one is stressed or not does probably change the answers since less time can be spent on answering or reflecting on the questions so I would guess that answers are less accurate in the afternoon when one wants to go home from work and relax.”
- **P3:** “Yes, I believe it had an impact, as for example, during the morning hours I was more eager to mark higher for positive emotions.”
- **P4:** “I think that it is likely that I filled the questionnaires differently compared to earlier times in the day.”
- **P5:** “More in a sense of experiencing different feelings at different times of the day, but not that I rated the survey differently at different times e.g. more positively because it was morning.”

- **P6:** “I'm not sure regarding the question. If it's about if the time of day would have made me give unprecise answers, then: No, I don't think so. If it's about if I generally saw that at given times of days I was more balanced/frustrated/stressed etc or if I felt more disturbed when filling out the survey, then: Yes.”
- **P7:** “I think that when answering questions at the end of the day there is probably a bias due to the fact that it is hard to remember how I felt at a certain point in the day. However, when answering questions on the iPhone or apple watch it didn't influence my answers as I was able to answer straight away how I felt at that time.”
- **P8:** “I don't think so. Maybe, I felt more stressed in the afternoon when my workday came to a close and I wasn't done with my work for today yet. So, in the afternoon surveys I might have felt more stressed than in the morning.”
- **P9:** “I was more annoyed filling out the surveys in the late evenings, so the computer surveys were maybe influenced a little bit.”
- **P10:** “Yes, I think so. In the morning, I feel more mentally fit, rested, and balanced. Throughout the day these qualities decrease due to the work I do.”
- **P11:** “Hard to say, but the email one (recalling my feeling at three different time on the same day) is surely very difficult.”
- **P12:** “In one sense, yes, because your feelings depend on the things happening and people being around you, so the more time of a day passes the more could have happened to influence the feelings. However, in general I cannot say the later I completed the questionnaire the worse were my answers. It could have been in both ways negative-> positive and vice versa.”

The results of the overall questionnaire show that eight out of the twelve participants think that the time of the day they completed the questionnaire did influence their answers in general. The other four participants have responded that the time of day did not influence their answers.

Additionally, participants were asked more specifically whether they think responding in the moment had an effect on their answer versus responding at the end of the day. The feedback regarding responding in the moment can be found in the first paragraph and the feedback regarding responding at the end of the day is described in the second paragraph.

Comments regarding “In the moment”

- **P1:** “No. I would come up with the same answers later.”
- **P2:** “The answers are more accurate.”

- **P3:** “I do not think so, however it was much easier to answer as it did not require recalling anything.”
- **P4:** “More erroneous and more spontaneous than in the evening. As I did the questionnaires 'on-the-go', I pressed the wrong button and could not go back. But in all, I did not think long when pressing the button on the phone and watch. I expect the inputs more volatile (or higher in variation).”
- **P5:** “No. I did not felt this.”
- **P6:** “I wouldn't say so. I tried to answer the questions independent from the modality.”
- **P7:** “No, I don't think so. I had no incentives to misreport my true feelings at time T.”
- **P8:** “No, I could give honest answers whenever I filled in the survey. However, I missed a few surveys because I didn't have time or didn't see the notification on my phone right away in that moment.”
- **P9:** “I think it was more accurate.”
- **P10:** “I already answered this in an earlier question.”
- **P11:** „Not really.“
- **P12:** “yes it depended on the situation and the influencing factors like workload, deadlines, private things to take care of.”

Comments regarding “End of the day”

- **P1:** “No. I would give the same answers earlier as in the evening.”
- **P2:** “The answers are less accurate. Thinking about parts of the day which I only vaguely remember made me a bit aggressive, that’s why I really disliked the computer questionnaire since I had the feeling that my answers are not accurate at all and that filling in the form is a bit a waste of time.”
- **P3:** “yes, as I was very often tired which could result in marking negative emotions higher, even for the morning hours.”
- **P4:** “The answers were more in the middle as I think that I thought more about my answers (and changed them during the filling in). I expect less variation when filling out the computer survey.”
- **P5:** “Yes, I think they were less accurate because it is more difficult to think back to a certain time and think about my feelings.”
- **P6:** “I don't think so. Even though I liked the computer-based surveys the least, I tried to give accurate answers.”

- **P7:** “Probably as once the day has passed it's sometimes difficult to think back about my true feelings. Or sometimes the day could have started in a very negative way but then end in a positive way and in the end when asked how I felt in the morning I would report weaker negative feelings than it truly was.”
- **P8:** “I don't think so. Sometimes I didn't remember exactly how I felt in the morning maybe.”
- **P9:** “It was sometimes hard to remember how I was feeling throughout the day.”
- **P10:** “I already answered this in an earlier question.”
- **P11:** “Yes, because I cannot recall my exact emotion.”
- **P12:** “Yes, because I could distance myself more from what was happening all day long. Often in the evening I could answer in calm situation as I was on my way home or at home.”

The comments above have been clustered and the result can be found in Table 17. Half of the participants felt that responding in the moment lead to more accurate measures and it was easier for them, because they did not have to recall their emotions. The other half of the participants reported that responding in the moment had no effect on their responses. Asking about the influence of responding at the end of the day, 10 out of 12 participants state that they believe responding at the end of the day had an effect on their answers. For the following reasons, being tired and therefore marking negative emotions higher, not remembering exactly how they felt during the day and taking more time to think about their answers in the computer questionnaire.

Do you feel that completing the questionnaire in the moment/at the end of the day influenced your answers? Please specify why and how.		
	Yes	**No**
In the moment	6	6
At the end of the day	10	2

Table 17: Results to In the Moment Compared to End of Day

Additionally, respondents were asked which time of the day they felt it most disruptive to fill in the survey on the smartphone and smartwatch. The results can be found in Table 18, where the number of times a person selected and not selected the answer option has been summed up.

Which time of the day did you find it most disruptive?				
	Smartphone Questionnaire		**Smartwatch Questionnaire**	
	Yes	**No**	**Yes**	**No**
Early in the morning	3	**9**	**4**	8
Late in the morning	**6**	6	3	9
Right before lunchtime	**5**	7	2	10
During lunchtime	3	**9**	2	10
Right after lunchtime	4	8	1	**11**
Early in the afternoon	4	8	2	10
Late in the afternoon	4	8	**4**	8
In the evening	3	**9**	3	9

Table 18: Results of the Influence of the Time of Day

For the smartphone questionnaire, the options of *Late in the morning* and *Right before lunchtime* have been selected the most as the times when it was most disruptive. The smartwatch questionnaire influenced participants the most *Early in the morning* and *Late in the afternoon*. The participants felt that the smartphone questionnaire was the least disruptive *Early in the morning*, *During lunchtime* and *In the evening*. The smartwatch questionnaire was perceived as the least disruptive *Right after lunchtime*.

5.4. Intrusiveness: Ease of Answering

The factor ease of answering has been qualitatively measured with questions regarding the accuracy of the duration of the questionnaire, the general burden to fill in the survey, the burden to fill in the survey five times per day, the difficultness to understand the questions, the surveys easiness and the efficiency to fill in the survey (see Chapter 4.2).

5.4.1. Accuracy of the Duration of Questionnaire

On a scale from 1 to 5, do you think the time necessary to complete a single survey was accurate? 1= not at all, 5 = very		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	37	3
Smartphone Questionnaire	52	5
Smartwatch Questionnaire	47	4

Table 19: Results of Accuracy of the Duration of Questionnaire

Participants perceived as the time to complete a single smartphone survey as the most accurate. Four out of six comments about the smartwatch questionnaire included that the number of

questions as well as the simplicity of answers made the survey quick. However, two participants stated that the survey was rather short to capture all feelings. The time to fill in a smartphone questionnaire has been perceived as accurate as well, but less accurate than the smartphone survey. The time to complete a computer questionnaire has been perceived as the least accurate and only moderately. The feedback about the computer questionnaire included that the survey did not require much time and the time estimation was too high.

5.4.2. Burden

On a scale from 1 to 5, how burdensome was it to fill out the survey? 1 = not burdensome at all, 5 = very burdensome		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	24	2
Smartphone Questionnaire	37	3
Smartwatch Questionnaire	23	2
Overall Questionnaire Rating	27	2

Table 20: Results Burden

The smartphone questionnaire has been perceived as moderately burdensome. Compared to the computer and smartwatch questionnaire, which are both a little burdensome, the smartphone questionnaire is more burdensome. The most mentioned negative point about the smartphone questionnaire was that they normally do not use their smartphone that often and only for communication purposes. Five participants commented on the computer questionnaire that it took more time to answer and the aspect of "going-back-to-work" was more burdensome. Therefore, they were also more annoyed by this type of questionnaire. Most participants commented that the smartwatch survey only takes a few seconds, and, in their opinion, it was more comfortable than the smartphone survey.

On a scale from 1 to 5, do you think filling out the questionnaire 5 times per day on the iPhone or Apple Watch was burdensome? 1 = not at all, 5 = totally		
Survey Type	**Sum**	**Rating**
Smartphone Questionnaire	22	2
Smartwatch Questionnaire	40	3

Table 21: Results Burden of 5 Times per Day

Participants stated that filling out the questionnaire 5 times per day on the smartwatch was more burdensome than on their smartphone. The reasons are the small screen on the Apple Watch and that when they were asked about their feeling's participants took the time to think and this

lead to a disrupted work flow, which furthermore did no please them. Additionally, one participant stated that even though the smartphone and smartwatch survey were quick and simple, they required constant checking for pop ups.

The feedback in the overall questionnaire has shown that all participants felt filling out the surveys a little burdensome. In general, some participants felt that filling out the survey was fairly simple, did not take a lot of time and the number of questions is appropriate to not interrupt the task or activity too much. One participant mentioned specifically that the questions should be filled with a person's gut feeling and, in the moment, to achieve the most accurate answers. Other participants mentioned that filling out the smartwatch and smartphone surveys gave them the feeling of needing to constantly check for notifications and the interruption was irritating for them.

5.4.3. Understanding of Questions

On a scale from 1 to 5, do you think the questions were difficult to understand? 1 = not at all, 5 = very		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	18	1
Smartphone Questionnaire	20	1
Smartwatch Questionnaire	17	1
Overall Questionnaire Rating	20	1

Table 22: Results Understanding of Questions

The questions of all three surveys have been rated as not at all difficult to understand. One participant mentioned that he thought these questions were straightforward. However various participants mentioned that the questions were phrased very similar and many feelings were opposites. Therefore, it was hard for them to distinguish between the feelings.

5.4.4. Easiness

On a scale from 1 to 5, how easy was it to complete the survey? 1 = not at all, 5 = very		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	51	5
Smartphone Questionnaire	56	5
Smartwatch Questionnaire	52	5
Overall Questionnaire Rating	53	5

Table 23: Results Easiness

All three survey types have been rated as very easy to be completed. Two participants mentioned that the computer questionnaire was not so easy to be completed, because it was hard for them to decide between the seven possibilities to rate their feelings. Half of the participants added as comment to their rating that the smartphone questionnaire was very easy, and the structure seemed well prepared. The smartwatch questionnaire received the same comments, however three participants mentioned that the buttons were very small and therefore, they were not sure if they have tapped the correct answer. The overall feedback questionnaire has shown that in general all three types were well designed and easy to fill out.

5.4.5. Efficiency

The factor efficiency of the participants is measured in terms of the time they needed to fill in the different survey types. Figure 12 displays the average time over all participants per day and per device. We can see that for all devices the response time decreased and therefore the efficiency gradually improved over the course of the study. The computer survey had an average response time between 171 seconds and 379 seconds. On the first attempt, participants needed on average 339 seconds to fill in the computer survey, which decreased by 19.7% to 272 seconds on average in their last attempt.

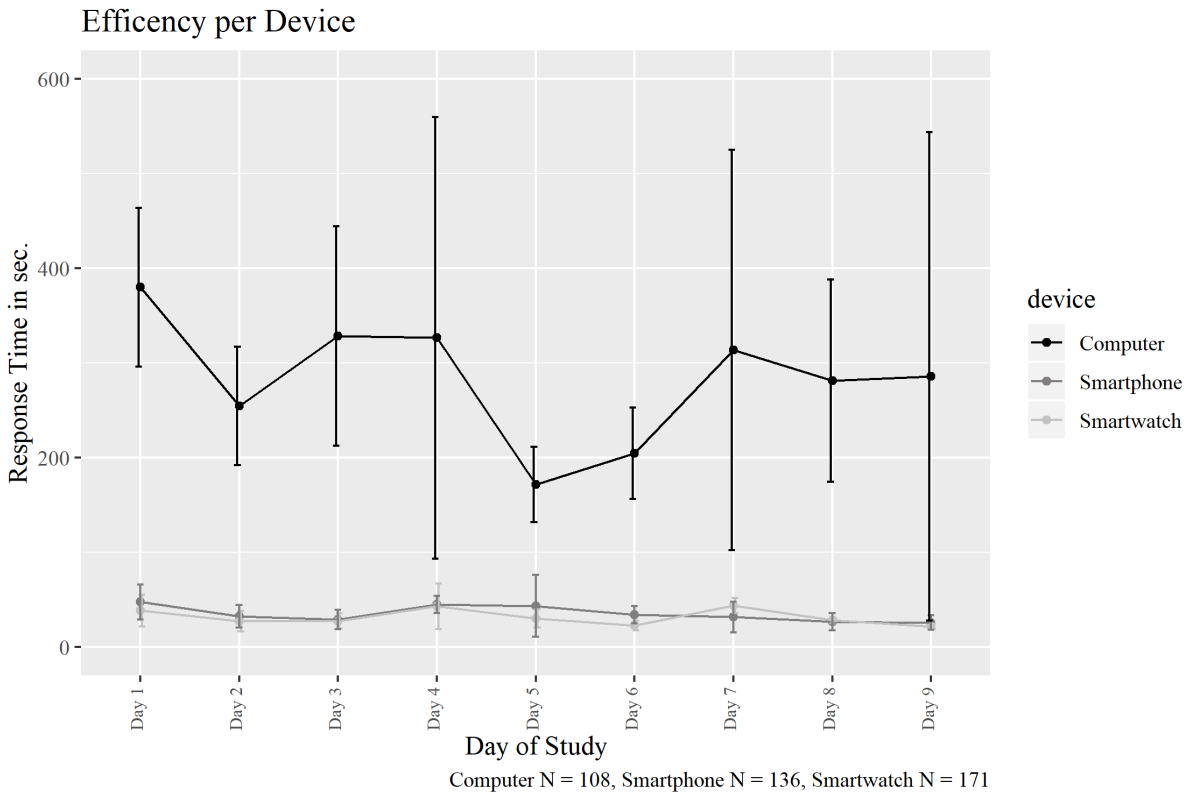


Figure 12: Average Response Time per Device per Day

The average response time on the smartphone ranged from 25 seconds to 71 seconds. The participants needed on average 49 seconds in their first attempt and on average 29 seconds in their last attempt, which results in a decrease by 40 %. The average response time on the smartwatch ranged from 27 seconds to 50 seconds. Additionally, it took the participants on average 41 seconds on their first attempt and 28 seconds on their last attempt to fill in the assessments. This results in a decrease by 32% for the smartwatch response time.

The average response time per device over the whole study is the smallest for the smartwatch with 33 seconds (see Figure 13). The smartphone survey had an average response time of 37 seconds and the computer survey an average response time of 282 seconds. Considering the

average response time per device and the decrease of the response time from the first to the last attempt, participants were more efficient on the smartwatch and the smartphone. However, which of these two devices types was more efficient, couldn't be assessed with this analysis.

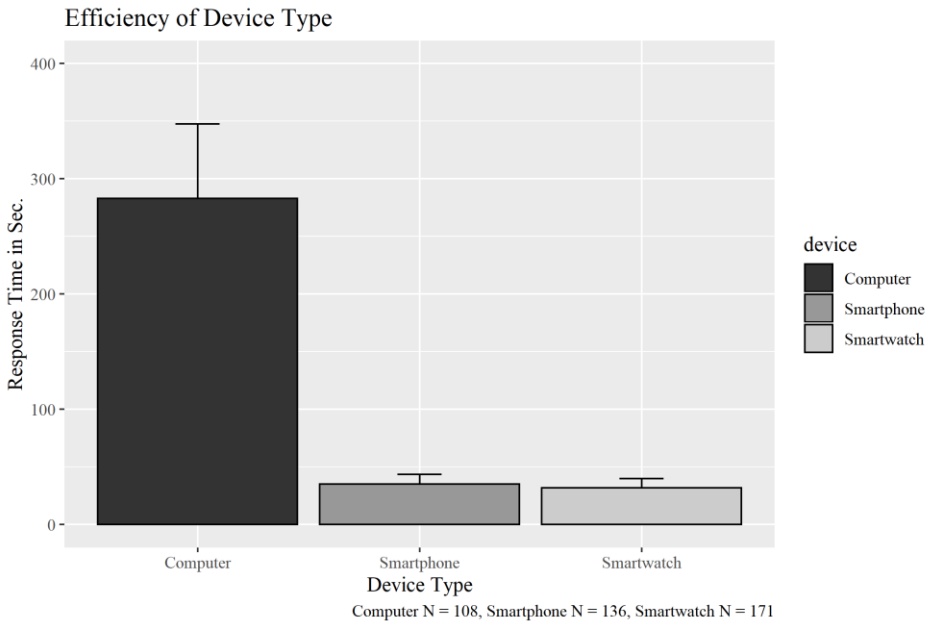


Figure 13: Average Response Time per Device

5.5. Perceived Accuracy

The accuracy of the feelings has been qualitatively assessed with questions regarding the influence of the question on the answer, a subjective assessment of the accuracy of the response, the way participants have been asked about their feelings, whether the week the study took place was representative or not for them and the reasons for not filling in an assessment (see Chapter 4.2).

5.5.1. Influence of Question on Answer

Did filling out the questionnaire influence how you felt? Please specify why.		
Survey Type	**Sum**	
	Yes	**No**
Computer Questionnaire	4	8
Smartphone Questionnaire	5	7
Smartwatch Questionnaire	5	7

Table 24: Results of the Influence of Question on Answer

Four participants answered that the questions in the computer questionnaire did influence how they felt, and eight participants answered it did not influence their feelings. The questionnaire had for those four participants the effect that it made them reflect about their feelings, which further gave them a good feeling about what they did and achieved that day. One participant stated that the survey had a negative effect on their feeling, because this person realized that he had marked overall negative emotions during the whole day. For both the smartphone and smartwatch questionnaire, seven participants answered this question with “No” and 5 participants responded with “Yes”. Again, those participants which answered this question with yes, commented that the survey made them think more about their mental state and about the reasons for their feelings. The participants, which answered this question with no described that their feelings stayed the same, their mood occurred due to external issues, such as deadlines, and was independent from the questionnaire.

5.5.2. Accuracy of Responses

On a scale from 1 to 5, how accurately do you think your responses reflect your feelings? 1= not at all, 5 = very		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	40	3
Smartphone Questionnaire	43	4
Smartwatch Questionnaire	45	4
Overall Rating	44	4

Table 25: Results Accuracy of Responses

The computer questionnaire was perceived as moderately accurate in reflecting the feelings. On the other hand, the smartphone and smartwatch questionnaires have been rated as more accurate than the computer questionnaire. Various participants stated that their mood in the evening influenced their entries for the different time slots. Additionally, participants felt that the time difference affected their accuracy on their responses. Regarding the smartphone questionnaire, participants only mentioned that their accuracy might have been affected by the fact that they probably made errors when selecting answers and that they tended to not select the extreme answer. Participants felt that the smartwatch questionnaire and smartphone questionnaire improved the accuracy of their responses. Furthermore, various participants mentioned in the overall feedback questionnaire that the set of items was limited, and it was difficult for them to really know how they felt. One participant mentioned specifically that in his opinion it depends on the survey and that he felt that the smartwatch and smartphone survey were more accurate than the computer survey. Other participants think the questions were accurate and tried to distinguish between the feelings and answer in a reasonable time. In general, participants felt it difficult to assess their current true feeling, for example whether they are truly stressed or if they feel balanced.

5.5.3. Way of Asking about Feelings

On a scale from 1 to 5, how much did the way you were asked about your feelings (i.e. the phrasing of the questions) influence the way you answered these questions? 1 = not at all, 5 = very much		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	22	2
Smartphone Questionnaire	25	2
Smartwatch Questionnaire	30	2
Overall Rating	25	2

Table 26: Results Way of Asking

The influence of the way of asking about the feelings is a little for all survey types. Multiple participants stated that they got used to the phrasing of the questions and answers and it seemed straightforward for them. Therefore, the participants think the way they were asked about their feelings did only influence their answers a little. Some participants mentioned that they gave a second thought to the questions as they did not want to exaggerate on their feelings. One participant mentioned that he was rather affected by the order of the feelings than by the way he was asked about them.

5.5.4. Representative Week

On a scale from 1 to 5, were these representative 9 days of your life? 1 = not at all, 5 = very	
Rating	**Number of participants**
1	0
2	2
3	1
4	4
5	5
Average Rating	**4**

Table 27: Results Representative Week

Most participants stated that these nine study days were very representative compared to their normal life. Two participants mentioned specifically that they were studying for exams and therefore were more emotionally and stressed than usual.

5.5.5. Missing Response

The participants have been asked for the reason why they missed a self-assessment. The results of this question are listed below.

- **P1:** “The phone questionnaire didn't work in the first day and didn't get the notifications. I got some notifications during my meetings and was not able to fill out the questionnaires.”
- **P2:** “Usually I could not respond when I had to fulfill important tasks like driving a boat or not filling in the survey during a meeting since this is bad meeting etiquette.”
- **P3:** “Trying to be as honest as possible, sometimes I realised I marked quite opposite emotions like being balanced and anxious with the same number. It made me think about how I felt and sometimes resulted in slight changes to the survey (only the computer one). I believe the reason was trying to hold off the negative emotions, especially during the morning hours, which could influence how I feel during the rest of the day.”
- **P4:** “had a small number of freezes (no submits) internet connection was not available - not sure whether it submitted the answers later”
- **P5:** “Mainly because I was talking to someone and thought I will answer it later to not disrupt the conversation. This sometimes led to the problem, that I forgot to answer it at all. A typical situation was a meeting with someone or also just a small coffee break. Another reason was, that I got the notification while I was on my way or catching a bus and therefore could not answer it.”
- **P6:** “I don't think I was unable to respond at any time. Sometimes it just took me longer to figure out how I feel in that given moment. :)”
- **P7:** “Once I missed the notification as I woke up too late (shouldn't ask questions over the weekend I think)”
- **P8:** “I had a busy weekend where I was out and about with friends. When I'm in company I don't check my phone so often or at all, which means I missed some of the iPhone surveys on that weekend.”
- **P9:** “Often I was not aware there was a new survey available, because it did not vibrate. This was especially the case on the watch, but I'm sure that will be fixed in future studies. If I was in meeting or eating lunch then I would fill out the surveys afterwards.”
- **P10:** “Either I was still sleeping, e.g. Sunday morning, or I was unavailable because of a conversation which could either be a meeting or talking to someone on the phone, etc.”
- **P11:** “Didn’t notice the notification”

- **P12:** “Sometimes I did not respond immediately, so that was because of having meetings at work or not seeing that a new survey was opened”

The responses of the participants have been grouped and aggregated in Table 28 to get a better overview of the reasons why participants did not respond and fill out a self-assessment.

If you were sometimes unable to respond, what were the main reasons?	
Company of friends or family	Did not notice notification
Woke up too late (at the weekend)	Postponed response and forgot later
Meetings	Trying to catch bus
Phone call	Internet connection problems
Application did freeze	Eating lunch
Fulfilling important tasks	

Table 28: Aggregated Results Missing a Response

5.6. Effect Device Type

On a scale from 1 to 5, did the form of the survey, i.e. that you answered the questions on your computer, iPhone or Apple Watch, influence how you answered the questions? 1= not at all, 5 = totally		
Survey Type	**Sum**	**Rating**
Computer Questionnaire	27	2
Smartphone Questionnaire	21	1
Smartwatch Questionnaire	25	2

Table 29: Results Input Type

The effect of the computer and smartwatch has been rated a 2 (i.e. a little) and the effect of the smartphone has been rated a 1 (i.e. not at all) on a scale from 1 to 5. However, various participants have stated that each type affected their answering behaviour in different positive and negative ways. The computer questionnaire had the effect that participants were less spontaneous and adjusted their answers by considering the previous answers. Making the input on the smartphone had the effect that some participants thought it was easier and therefore they were less annoyed. The smartwatch questionnaire had the effect that two participants believed they answered with a tendency towards the extremes.

Analysing the actual ratings of the stress level gathered during the study regarding the effect of the device type on the stress level does not give any clear results (see Figure 14). The average stress ratings do disperse a lot between the participants, and it is therefore no clear trend identifiable. However, the stress rating gathered with the computer and smartphone seem to be more stable during the nine days of study compared to the smartwatch ratings.

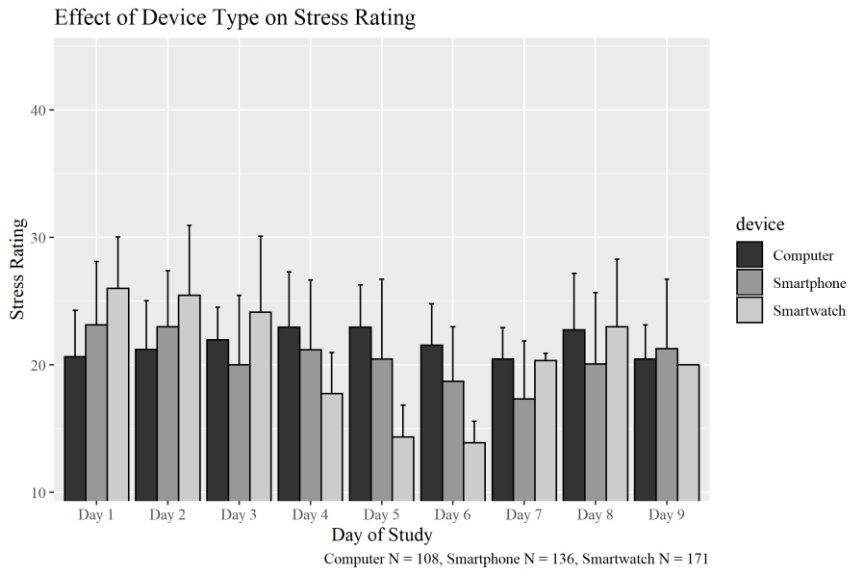


Figure 14: Effect of Device Type on Stress Rating per Day

6. Discussion

A field study has been conducted to answer the two research questions: *RQ1: How do different factors affect the user's experience of the data collection?* and *RQ2: How can self-reported data about stress be gathered to yield a positive experience for the user?* The results (see Chapter 5) have been summarized in Table 30 and will be discussed in this section with regard to the formulated hypotheses and to the related work in this research field.

Measure	**Computer Survey**	**Smartphone Survey**	**Smartwatch Survey**	**Scale**
Comfort				
Ranking	3	2	1	1 = best 3 = worst
Usability	5	5	5	1= not at all 5 = totally
Naturalness	No	Yes	Yes	Yes or no
Design	Moderate	Good	Good	Bad to Very Good
Intrusiveness				
Disruptiveness	2	3	2	1= not at all 5 = very
Environmental Influence	2	2	2	1= not at all 5 = a lot
Timing of self-report: in the moment vs. end of the day	Yes	Yes/No	Yes/No	Yes or No
Timing of self-reports: specific times		Late in the morning Right before lunchtime	Early in the morning Late in the afternoon	
Accuracy of duration of Questionnaire	3	5	4	1= not at all 5 = very
Burden	2	3	2	1 = not at all 5 = very
Burden 5 times		2	3	1 = not at all 5 = totally
Understanding of Questions	1	1	1	1 = not at all 5 = very
Easiness	5	5	5	1 = not at all 5 = very
Efficiency	3	2	1	1 = best 3 = worst
Perceived Accuracy				
Influence Question on Answer	No	No	No	Yes or No
Accuracy of response	3	4	4	1= not at all 5 = very
Way of asking about feelings	2	2	2	1 = not at all 5 = very
Representative Week	4			1 = not at all 5 = very
Effect Device Type				
Effect Device Type	2	1	2	1= not at all 5 = totally

Table 30: Summarized Results

6.1. Discussion Research Question 1

How do different factors affect the user's experience of the data collection?

The results show that two different categories of influencing factors can be identified. The first category are factors, which were perceived as having the same effect on the user's experience no matter the survey type (see Table 30 dark grey background). The second category contains the factors, which were not perceived as having the same effect on the user's experience for the different survey types (see Table 30 white background). The cells with black background in Table 30 represent the questions, which could not be asked for the computer questionnaire due to its different methodological approach.

The first category covers the factors usability, environmental influence, understanding the questions, easiness, influence of the questions on the answer and the way of asking (for details see Chapter 4.2 & 5). We assume that the usability of the different surveys did not affect the user's experience in our study, because the design and structure of the surveys enabled the participants to understand immediately how they could enter their responses. Furthermore, we used existing design principles to make the questionnaires more intuitive for the user and to enable him to apply his current interaction knowledge. The reason why the environment does influence the user in the same way for all different survey types is that in general participants answer slightly different in the presence of others or try to hide the questionnaire from others. The feedback has shown that from a user's perspective this does not depend on the survey type, but rather on the fact whether his or her environment is part of the emotions or not. Moreover, participants did not perceive that the survey type had an effect on the difficulty to understand the questions. We suppose that the participants were able to suppress the influence of the device type by focusing on the content, rather than on the user interface and the interaction itself. The factor easiness has been equal across all survey types, which means no matter the survey type it was easy for participants to respond. We believe that this has been supported by the usability and the difficulty to understand the question. Additionally, participants did not perceive that the device type and method had an effect on whether the questionnaire influenced how they felt. Participants described that the surveys in general made them think more often about their mental state than usually and this was again not linked to a specific survey type. Finally, the way the participants were asked about their feelings has not been influenced by the method or device type. In contrary, participants mentioned that they got used to the phrasing of the questions and they understood them right away. Once again, the participants were able to blend out that they were answering on different devices.

The second category includes the factors ranking, naturalness, design, disruptiveness, timing of the self-reports, accuracy of the duration of the questionnaire, burden, efficiency, perceived accuracy of the responses and the effect of the device type (for details see Chapter 4.2 & 5). These factors are either positively or negatively connected with the user's experience of the data collection depending on the survey type. Connecting our results to the related research has shown that some findings could be validated whereas others not. For example, Atz [1] claims that the researcher should rather minimise the complexity of the input (amount and type of questions, interface and integration) than reduce the number of prompts. However, this could not be verified with our results. They rather suggest that perceiving the number of prompts as burden is connected to the device type and the self-reports themselves, than to the complexity of the input. The main reasons why a survey type has been perceived as burden were the need to get the device and the survey itself, because it made people think more about their emotions, which is time intensive and interrupts their workflow. These results are supported by Hernandez et al. [33], who have found that minimizing the time needed to access the device is critical to ensure the highest quality of the response and decrease the burden.

The different factors and their effect on the user's experience will be discussed in more detail in relation to the hypotheses of this thesis in the next few paragraphs. An overview over the factors which do have an effect, has demonstrated that the perceived comfort and enjoyableness seem to be connected to the perceived intrusiveness of the device. In other words, if the device and method to gather the self-reported data of stress do not comfort the user, it is perceived as more intrusive. Furthermore, the perceived accuracy of the method is depending on the perceived intrusiveness of the device. If the methodological approach and device are perceived as intrusive, it is also perceived as it does not lead to accurate measures. Moreover, the feedback has verified that participants prefer to achieve accurate measures and to use non-intrusive types of data collection.

6.2. Discussion Research Question 2

How can self-reported data about stress be gathered to yield a positive experience for the user?

In order to answer the second research question, the four associated hypotheses (see Chapter 1) will be discussed individually and a final statement regarding this research question will be made in Chapter 6.3.

H1: The perceived most comfortable way to collect self-reported data about stress is with an ESM approach on a smartwatch.

The ranking question has shown that participants preferred the smartwatch survey the most. The participants described that it was much faster and more convenient than the smartphone and computer questionnaire. Moreover, with its handiness and the slight vibration it was perceived as less intrusive than the two other conditions. Regarding the ease of usability participants specified that the survey description, the study schedule as well as the survey content were well structured and easy understand (cf. P3). Therefore, all three survey types achieved the same results regarding the usability. Additionally, participants felt that responding in the moment was more natural and easier to report their true feelings (cf. P7). Moreover, they preferred the smartphone and smartwatch design rather than the computer survey.

Overall, we found that the smartwatch received better ratings in terms of comfort, enjoyableness, usability, naturalness and design (see Table 30). Hernandez et al. [33] reached to the same conclusion in their study on ESM with wearable devices. Two participants in the About your Day study mentioned that filling out the survey on the smartwatch unexpectedly added some value for them. It led to the fact that their friends and colleagues were interested in knowing what they were doing on their smartwatch. In addition, the participants mentioned that they finally had something more interactive to do with the smartwatch, whereas it is usually just displaying information (cf. P2, P10). The second most comfortable way is the smartphone survey, which lacks regarding the comfort and enjoyableness of the participants. This has also been reflected in their comments, which describe that they perceived the smartphone as more distracting and in presence of others it was more unpolite to use it. The least comfortable way is the computer survey, which participants did not perceive as natural and lacks regarding the design aspects. Furthermore, participants gave it the lowest ranking regarding the comfort and enjoyableness of the survey. This was also reflected by their comments, which described that participants thought it was annoying, too distant and it was too hard to reconstruct their day. This leads to the fact that hypothesis 1 can be verified and the most comfortable way to collect self-reported data about stress is with an ESM approach on a smartwatch.

H2: The perceived least intrusive way to collect self-reported data about stress is with an ESM approach on a smartwatch.

The results of the disruptiveness question have shown that participants perceived the smartphone questionnaire as more disruptive compared to the computer and smartwatch

questionnaire. This has partly been reflected in the comments of the participants, which have mentioned for example that it was little more disruptive than the watch, but less compared to the computer version (cf. P10). Various participants have justified their reasoning and rating with the argument that while they were working, they always set their phone to silent in order to not interrupt their colleagues. Besides, in their opinion it was more socially accepted to do something on their smartwatch than doing something on their smartphone in the presence of others. The smartwatch and computer questionnaire have both been rated as a little disruptive, but for different reasons. Participants mentioned that the smartwatch interrupted their workflow and some of the participants normally do not wear their smartwatch on the weekends. For the computer survey, it was most disruptive for the participants to get their laptop. Moreover, the fact that filling in something on their laptop has given them the impression to be back at work, while they were relaxing at home. The influence of the environment on the answers of the self-reports has been rated as a little for all three survey types. This has also been reflected in the comments of the participants, which described that they tend to answer more positively in presence of others. However, one participant mentioned that it did not affect the answer, because the people in his environment were not able to see what he was answering on his smartphone or smartwatch (cf. P5). Nearly all participants agreed that the timing of the self-reports did influence their answers. Most participants mentioned that especially filling out the questionnaire at the end of the day did influence their answers, because it was hard to remember how they felt at a certain point during the day (cf. P7). Additionally, one participant described that his answers in the computer questionnaire were more in the middle of the scale, as he tended to think more about his answers (cf. P4). Furthermore, most of the participants felt that responding in the moment did influence their answers, because they did not have to recall feeling a certain emotion throughout the day. Hence, their answers were perceived as more accurate. The ratings have demonstrated that participants perceived the duration of the questionnaire was most accurate for the smartphone survey. The duration of the smartwatch questionnaire was perceived as the second-most accurate, whereas the duration of the computer questionnaire was perceived as the least accurate. However, considering the comments about the computer and smartphone survey has indicated that not all participants understood the question. Most participants compared the time they actually needed to fill in the survey to the estimated time by the examiner. Instead of their perception if the time they needed to fill in the survey was accurate for a survey like this. Therefore, the computer questionnaire received a low rating regarding the accurateness of the duration, because from the participant's point of

view the estimation was not accurate and they were much faster in filling in the survey. The number of questions as well as the simplicity of the answers made the smartphone and smartwatch survey quick, which was appreciated by various participants. However, one participant mentioned that in his opinion the survey was very short to capture the feelings (cf. P4). The smartphone survey has been rated as being more burdensome to fill in compared to the computer and smartwatch questionnaire. The reason why participants perceived the smartphone survey as burdensome was in most cases that they felt pressure to check their smartphone more frequently than usual (cf. P3, P4, P8). However, for all three survey types the ratings have been very disperse from 1 to 5, which is also reflected in the comments of the participants. Some of them have mentioned that as they knew when they would approximately receive the invitation to fill in the computer questionnaire it was easy for them to incorporate it in their daily schedule (cf. P3). In addition, the survey does not require a lot of cognitive load (cf. P10), which made it not very burdensome for the participants. Others stated that the computer questionnaire was very burdensome, because they had to get their laptop out of their bag, sit down, fill in the survey and put the laptop back (cf. P6, P9). In another case a participant rated the smartwatch survey as burdensome due to the fact that the participant felt it was a lot of reporting and that he thought he would not have the discipline to keep that frequency of reporting up for a long time in normal life (cf. P8). Asking specifically about the burden of answering 5 times per day has indicated that participants think it is more burdensome on the smartwatch than on the smartphone. One participant mentioned that filling in the survey on the smartphone was for him less cumbersome than on the smartwatch, because the interface was faster and the buttons bigger (cf. P8). All participants agreed that it was not difficult at all to understand the questions in each of the survey types. However, most of them stated that the different feelings were very similar, and it was hard for them to distinguish between them (cf. P2, 12). Additionally, one participant explained that in his opinion many feelings were opposites and if one of them is being asked, there is no need to ask the second one too (cf. P9). All survey types have been rated as very easy to be completed, which is also confirmed in the comments of the participants that the questions and structure seemed well prepared (cf. P3, P4, P6). Several participants have described that it was hard for them to decide between the options in the 7-point Likert scale question in the computer questionnaire. Furthermore, they suggest having the three consideration periods of the day (morning, midday, afternoon) next to each other, such that participants could see how their mood changed over their day (cf. P6). Considering the feedback of the participants regarding the computer questionnaire, this

suggestion would decrease the accuracy of the questionnaire, because participants would take more time to even out their ratings and be less spontaneous. Comparing the efficiency of the three survey types has shown that the average response time for the smartwatch survey was the smallest. Furthermore, the decrease in the average response time from the first to the last attempt was the highest for the smartphone.

Considering all the different factors described above, such as disruptiveness and burden, we conclude that the DRM computer survey is perceived as the most intrusive type. Comparing the ESM smartphone and ESM smartwatch results, we argue that the smartwatch survey performs better than the smartphone survey regarding the disruptiveness, burden and efficiency factor. Consequently, hypothesis 2 can be verified and the perceived least intrusive way to gather self-reported data about stress is with an ESM smartwatch survey. The results regarding the higher burden of the smartphone are also verified by Hernandez et al. [33], who have found that pulling the phone from the pocket or bag was less convenient than filling in the survey on the smartwatch.

H3: The ESM is perceived as the method that results in more accurate self-reporting measures compared to the DRM method.

Most of the participants responded that completing the questionnaire did not influence how they felt and therefore had no influence on their answers. This is also reflected by their individual comments which outline that their mood was independent from the questionnaire and that it did not trigger any kind of cognitive process that would influence how they felt (cf. P4, P8). Asking the same question in the smartphone and smartwatch feedback questionnaire has demonstrated that only four participants (P5, P8, P11, P12) did always respond the same, namely that filling in the questionnaire did not influence how they felt. The other participants did at least once respond to this question with yes, i.e. that completing the questionnaire had an influence on how they felt. The feedback has shown that especially with the computer questionnaire the probands started to reflect on what happened during the day. On one hand, this gave them a good feeling of what they had achieved that day (cf. P6, P9, P10). On the other hand, when the participants realised that they marked specific negative emotions as being stronger during the whole day, the computer survey made them feel slightly miserable (cf. P3). For another participant the survey had the negative effect that reading the word *stressed* or *in a hurry* made him start to feel that way (cf. P1). Additionally, one participant wrote that the questions forced him to think about himself in times he did not wanted to (cf. P7). The results of the accuracy of

responses have pointed out that responses are perceived as the least accurate in the computer questionnaire. The perceived accuracy for the smartphone and smartwatch questionnaire has been rated the same and more accurate than the computer questionnaire. This is also reflected in the comments of the participants. Most of the participants described that the DRM method is less accurate, because they were not sure how they felt in the morning and it was just a rough estimation on how they felt. For the smartphone and smartwatch survey, participants specified that they were not always sure if they accidently pushed the wrong button. Therefore, they were uncertain about the accuracy of their feelings. In general, most of the participants found it difficult to scale their feelings and to decide what the real feeling was, which might have an effect on the accuracy as well (cf. P1, P12). The way how participants were asked about their feelings has only a little effect in all survey types. This is also verified by their comments that it does have a small effect regarding the order of the feelings. For example, one participant mentioned that he once pressed a higher number for the feeling *stressed* and a lower number for the feeling *in a hurry*, because the question regarding the feeling *stressed* appeared first. In that sense the participant associated a higher value to that feeling and to equalize the answers he then selected a lower value for the second similar feeling (cf. P12). Another participant described that sometimes the order first covered all negative emotions and then all positive emotions (cf. P3). In this case, it was easier for the participant to first evaluate the negative emotions and then focus on the positive aspects. Therefore, the way of asking was perceived as having a little effect in all survey types.

Overall the results have shown that the ESM is perceived as resulting in more accurate measures than the DRM. Hence, hypothesis 3 can be verified. In the end, participants perceived the nine-days of study were a representative week for them. However, there might be some systematic weekend effects regarding the responding frequency and probably regarding the stress level. The main reasons for missing a survey were in most cases the presence of others or that the participants were working and not able to respond. In contrast to Hernandez et al. [33], we found no difference between the perceived accuracy of the smartphone and smartwatch survey. This might be due to technical advances in the last three years and/or using different brand products, Apple devices instead of Samsung devices. Moreover, the participants in our study were using their private devices instead of study devices as Hernandez et al. [33] did, which rules out the effect of novelty and might have an effect on the perceived accuracy.

H4: The reported stress level is perceived to be unaffected by the way how the self-reported stress data is collected.

The results have shown that participants think the smartphone survey does not have any effect on their answers. The computer and smartwatch questionnaire are perceived as having a little effect on how the participants responded in the survey. The feedback has shown that the influence of the smartphone has been rated as the smallest, because it was very easy to answers and people were less annoyed. The feedback regarding the computer questionnaire has shown that participants perceived their answers as less spontaneous. Additionally, some of them liked that they were able to change their answers in the computer questionnaire, which means it also influenced their answers. For the smartwatch questionnaire the feedback has pointed out that participants assume to have selected more extreme answers, due to the smaller scale. Considering that the smartphone survey made use of a 4-point Likert scale too, we are unable to determine whether the effect is certainly caused by the smartwatch survey. The analysis of the stress ratings has shown that they are very disperse for each survey type. The reasons might be due to different number of observations for each survey type and an effect of the assessment day on the perceived stress level. Therefore, we found no clear trend, whether a specific device type had an influence on the stress rating. Consequently, hypothesis 4 can neither be verified nor be falsified.

6.3. Discussion Summary and Recommendations

This section provides a suggestion on the way how self-reported data about stress should be gathered to lead to a positive experience for the user. From the author's point of view, the ESM smartwatch survey covers the way how self-reported data about stress should be gathered to yield a positive experience for the user. It outperforms the other survey types in the factors comfort, intrusiveness and perceived accuracy. This is also reflected by the received feedback, in which participants described that they felt the smartwatch survey was very lightweight and easy to do. Furthermore, many participants have mentioned that the length and amount of questions was just appropriate for the smartwatch. Some of the participants were even surprised by the fact that the application did work so smoothly, because they were not used to do a task on their smartwatch. In addition, it was convenient and natural for the participants to fill in the survey on their smartwatch during the day.

However, the smartphone survey is certainly a good alternative to the smartwatch survey. This has also been reflected by the comments of the participants. Various participants have described

that they normally do not wear their smartwatch all the time and therefore they would possibly miss some notifications. Additionally, they recognized that the UI on the iPhone is bigger, and the smartphone is faster. Participants were also uncertain if their response has been submitted on the smartwatch and if it is generally reliable. This has not been mentioned regarding the smartphone survey. The main problem with the smartphone is the negative connotation which is associated to it. Various participants have described that they feel uncomfortable and impolite to do a task on their smartphone in presence of other people. Considering the functional and interaction possibilities on the smartphone compared to the smartwatch, we propose to consider the smartphone as more appropriate in the context of stress measurement analysis and as intervention tool. From our perspective, combining an application, which gathers the self-reported stress data on the smartwatch and illustrates the analysis and trend on the smartphone, would overcome the described challenges with the smartwatch and smartphone. The application could even consider giving the possibility to gather the self-reports on the smartphone and meanwhile start recording the heart rate with the linked smartwatch. In order to improve the study design and application the next paragraphs summarize some recommendations worth being considered.

Recommendations Study Design

With regard to improve the study design, analysing the feedback of the participants has shown that from their point of view the feedback questionnaire did include repeating questions and the questions were too similar to distinguish between each other. The feedback has further revealed that participants would wish to have the activity options *eating* and *studying* in the questionnaire. Furthermore, participants missed the option *With random people* regarding the human environment, they are currently spending time with. Additionally, various participants outlined that the 6-point Likert scale in the computer questionnaire was too large to distinguish and assess their feelings. They preferred the 4-point Likert scale on the smartwatch and smartphone. In addition, two participants remarked that the positive values for the negative feelings (overwhelmed, stressed, in a hurry, mentally exhausted, overcommitted, tense, anxious, under time pressure) were the values 1 (Not at all) and 2 (Moderately) (cf. P2, P8). Whereas the positive values for the positive feelings (calm, rested, balanced) were the values 3 (Very) and 4 (Extremely), which was confusing for the participants. Hence, they suggest improving the scale for calm, rested and balanced. This would decrease the ease of usability by adding more cognitive load to the task, but it would improve the ability to distinguish between

the feelings. The study design could be enhanced by considering the factors, which received the same rating in all different surveys, as control variables. Finally, we assume a study with a longer duration could assess in-depth if the perceived level of stress is unaffected by the way how self-reported stress data is collected and rule out the effect of the assessment day.

Recommendations Application

Regarding the technical functionalities of the surveys, the participants were overall satisfied. However, one participant mentioned that it would improve the experience, if the answers could be saved locally and that they would be submitted as soon as the device connects to Wi-Fi. Some participants wished to have a back button in all survey types, in case they selected an option by mistake. However, we assume that this would decrease the accuracy of the measure, because users would adjust their ratings if they realise that they are reporting their feelings too honest. Moreover, it would also increase the time to fill in the survey. Considering that the duration of the smartphone and smartwatch questionnaire was highly appreciated by the participants, we would not recommend implementing a back button in the smartphone questionnaire and making the back button in the smartwatch questionnaire more obvious.
Additionally, we received some technical feedback regarding the notifications, which for example did not always trigger the watch to vibrate, although it was selected by the platform and the participant had selected it in the settings. This problem could not be reproduced and therefore, no improving proposal can be made. In general, participants did not encounter a lot of technical issues, besides the fact that the submission screen popped-up when the application was gathering the heart rate of the participants. However, they have been informed about this issue before the study. Overall the application did perform very well and only a few crashes have been reported in the feedback. As the cause of these crashes were not reported, hence, they cannot be assessed.

7. Conclusion

This thesis first summarized the theoretical and methodological foundations of emotions, which have shown that a variety of approaches exist to define and classify emotions. This variety is also reflected in the disagreement on conceptualizing emotions and interpreting their role in life [56]. Nevertheless, emotions are explored, and current research focuses on stress, which is of high importance due to its connectivity with burn-out. Stress is defined as a process by which a stimulus elicits an emotional, behavioural and/or physiological response [35]. Research indicates that not only physical, physiological and behavioural measurements achieve accuracy in assessing the level of stress, but also self-reports and the associated perceived level of stress are valuable. The most known self-reporting methods are the Experience Sampling Method (ESM) and Day Reconstruction Method (DRM). Drawing upon these findings, a within-subject study has been conducted to assess how different influencing factors affect the user experience of data collection and how self-reported stress data should be gathered to yield a positive experience for the user. The study compares the two methodological approaches ESM and DRM as well as the use of three different devices: computer, smartphone and smartwatch. The results suggest that there is no ultimate solution that fits all contexts and applications. This means that the different devices and methods evaluated in this study offer different benefits for different scenarios. However, focusing on the scenario of data collection of stress the ESM smartwatch survey received more positive scores across different influencing factors than the ESM smartphone survey and the DRM computer survey. The ESM smartphone survey is perceived to be more disruptive and the DRM computer survey is perceived to lead to inaccurate measures. However, combining the devices smartwatch and smartphone would overcome their specific drawbacks and improve the user's experience. In the end, the implemented application gives a lot of possibilities for extensions and configurations for future research.

8. Limitations

Our results are giving important insights into providing users with positive experiences while self-reporting of stress. Nevertheless, studies generally have limited resources and are subject to restrictions, which are outlined in the following. Our sample is limited to iPhone and Apple Watch users who encountered the opportunity to participate in the study, and who were pre-selected to match the given criteria. Hence, we assume that the results of the study implicate sampling and selection bias. In addition, our sample is limited to 12 participants, which implicates that the bias might be even larger. Additionally, people in a study are always more conscious and therefore feel the urge to commit and respond to as many self-assessments as possible. Moreover, a learning effect is associated with the applied type of questions, which leads to participants getting faster at reflecting and reporting over time. These are threats to the internal validity of the study and therefore threats to the extent to which the results support the claim of cause and effect.

Furthermore, we only tested a limited number of devices. Considering the variety of devices, our results may not apply to each of them. This implies that the results are not generalizable, but it would be interesting to validate the results with other available smart devices. Another limitation of this study results from the fact that when a questionnaire is being developed the researcher takes his own decisions and assumptions concerning the importance of questions into the development of the questionnaire. In addition, many researchers have pointed out that being repeatedly asked about particular feelings and behaviours, this may induce those feelings and behaviours or may cause participants to alter their behaviour [48]. This indicates that the perceived comfort, intrusiveness, accuracy and the perceived overall effect of the device type are also connected to the number of times a user is asked to fill in an assessment.

9. Future Research

In order to continue this research, next steps would be to improve the heart rate data gathering on the Apple Watch (see Section 3.2.4). Moreover, the available measurement values of the heart rate could be linked to the perceived level of stress measures from the About your Day study. The current heart rate implementation has proven to work well. However, the heart rate needs to be measured over a longer time period to provide insights into the stress level and therefore the implementation approach needs to be improved regarding the battery consumption. Advances in consumer wearables also provide the possibility to collect health data in a more extensive way than just the heart rate. Therefore, extending the prototype to record, for example, the respiration and galvanic skin response would provide more insights into the connection between the perceived stress level and the physiological measure of stress. In general, not all the collected data has been analysed and discussed in this thesis. Hence, some effort could be made to analyse the perceived level of stress in more in-depth and answer new research questions. In that sense, a new research question could be formulated as follows: how is the perceived level of stress connected to the activities the participant was performing and his demographic background? Additionally, all the recommendations regarding the study design and application should be implemented and tested in a next version (see Chapter 6). Another focus could be set on the exploration of creative input methods for self-reporting and run tests with different types of designs on the smartwatch (cf. [28]). Furthermore, technological developments support that research is not being limited to personal input devices anymore. Researchers can start to integrate data from external devices and networks to construct a richer context of participants [2]. Moreover, a long-term study could investigate our assumption that participants prefer to receive the behavioural intervention on their smartphone instead of their smartwatch (see Chapter 6).

Finally, a more extensive iteration of this intervention tool could focus on providing approaches to behavioural intervention (e.g. didactic lessons, exercises) or coaching according to the perceived stress level measures. Context-aware systems can sense a person’s behaviour and mental state and coupled with a treatment platform, it can positively reinforce adaptive behaviours and provide support for changing those [6]. This would contribute to learn dealing with mental health issues [6].

References

[1]: Atz, U. (2013). Evaluating Experience Sampling of Stress in a Single-Subject Research Design. *Personal and Ubiquitous Computing*, 17(4), 639–52. doi: 10.1007/s00779-012-0512-7.

[2]: Van Berkel, N., Ferreira, D. & Kostakos, V. (2017). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, 50(6), 93:1–93:40. doi: 10.1145/3123988.

[3]: Bolger, N., Davis, A. & Rafaeli, E. (2003). Diary Methods: Capturing Life as It Is Lived. *Annual Review of Psychology*, 54, 579–616. doi: 10.1146/annurev.psych.54.101601.145030.

[4]: Brüser, C., Stadlthanner, K., de Waele, S. & Leonhardt, S. (2011). Adaptive Beat-to-Beat Heart Rate Estimation in Ballistocardiograms. *IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society*, 15(5), 778–86. doi: 10.1109/TITB.2011.2128337.

[5]: Bundesamt für Statistik (2014). Publikation der Standardtabellen der Schweizerischen Gesundheitsbefragung 2012. Knapp ein Fünftel der Erwerbstätigen starken Belastungen im Beruf ausgesetzt. Medienmitteilung. Available at: https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/medienmitteilungen.assetdetail.37758.html [Accessed: 29.07.2019].

[6]: Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J. & Giangrande, E. (2011). Harnessing Context Sensing to Develop a Mobile Intervention for Depression. *Journal of Medical Internet Research*, 13(3). doi: 10.2196/jmir.1838.

[7]: Bylsma, L. M., Taylor-Clift, A. & Rottenberg, J. (2011). Emotional Reactivity to Daily Events in Major and Minor Depression. *Journal of Abnormal Psychology*, 120(1), 155– 67. doi: 10.1037/a0021662.

[8]: Carneiro, D., Novais, P., Augusto, J. C. & Payne, N. (2019). New Methods for Stress Assessment and Monitoring at the Workplace. *IEEE Transactions on Affective Computing*, 10(2), 237–54. doi: 10.1109/TAFFC.2017.2699633.

[9]: Chan, M., Estève, D., Fourniols, J.-Y., Escriba, C. & Campo, E. (2012). Smart Wearable Systems: Current Status and Future Challenges. *Artificial Intelligence in Medicine*, 56(3), 137–56. doi: 10.1016/j.artmed.2012.09.003.

[10]: Chan, Y., Ho-Wai So, S., Dun Ping Mak, A., Tien Ho Siah, K., Chan, W. & Wu, J. C. Y. (2019). The Temporal Relationship of Daily Life Stress, Emotions, and Bowel Symptoms in Irritable Bowel Syndrome-Diarrhea Subtype: A Smartphone-Based Experience Sampling Study. *Neurogastroenterology and Motility: The Official Journal of the European Gastrointestinal Motility Society,* 31(3). doi: 10.1111/nmo.13514.

[11]: Cohen, S., Kamarck, T. & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior,* 24(4), 385–96. doi: 10.2307/2136404.

[12]: Cohen, S., Kessler, R. C. & Underwood Gordon, L. (1997). Measuring Stress: A Guide for Health and Social Scientists. New York, NY: Oxford University Press.

[13]: Colombetti, G. (2009). From affect programs to dynamical discrete emotions. *Philosophical Psychology,* 22(4), 407–25. doi: 10.1080/09515080903153600.

[14]: Csikszentmihalyi, M., and Larson, R. (1983). The Experience-Sampling Method. *New Directions for Methodology of Social & Behavioral Science,* 15, 41-56.

[15]: Csikszentmihalyi, M., and Larson, R. (1987). Validity and Reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease,* 175(9), 526–36. doi: 10.1097/00005053-198709000-00004.

[16]: Csikszentmihalyi, M., Larson, R. & Prescott, S. (1977). The Ecology of Adolescent Activity and Experience. *Journal of Youth and Adolescence,* 6(3), 281–94. doi: 10.1007/BF02138940.

[17]: Daylio - The Big Idea of Micro Diary (2017). Available at: https://daylio.computerflow.io/ [Accessed: 21.07.2019].

[18]: DeVries, M. W., Delespaul, P. & Caes, C. I. M. (2001). The Experience Sampling Method in Stress and Anxiety Research. In Griez, E. J., Faravelli, D. N. & Zohar, J. (Eds.), *Anxiety Disorders* (pp. 289–306). John Wiley & Sons, Ltd. doi: 10.1002/0470846437.ch15.

[19]: DeVries, M. W. and Delespaul P. (1989). Time, Context, and Subjective Experiences in Schizophrenia. *Schizophrenia Bulletin,* 15(2), 233–44. doi: 10.1093/schbul/15.2.233.

[20]: Diener, E., and Tay, L. (2014). Review of the Day Reconstruction Method (DRM). *Social Indicators Research,* 116(1), 255–67. doi: 10.1007/s11205-013-0279-x.

[21]: Dockray, S., Grant, N., Stone, A. A., Kahneman, D., Wardle, J. & Steptoe, A. (2010). A Comparison of Affect Ratings Obtained with Ecological Momentary Assessment and the

Day Reconstruction Method. *Social indicators research*, 99(2), 269–83. doi: 10.1007/s11205-010-9578-7.

[22]: van Eck, M., Berkhof, H., Nicolson, N. & Sulon, J. (1996). The Effects of Perceived Stress, Traits, Mood States, and Stressful Daily Events on Salivary Cortisol. *Psychosomatic Medicine*, 58(5), 447–58. doi: 10.1097/00006842-199609000-00007.

[23]: Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. doi: 10.1080/02699939208411068.

[24]: Fliege, H., Rose, M., Arck, P., Levenstein, S. & Klapp, B. F. (2001). PSQ - Perceived Stress Questionnaire. In Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (Eds.), *Elektronisches Testarchiv (PSYNDEX Tests-Nr. 9004426)*. Trier: ZPID. doi: 10.23668/psycharchives.351.

[25]: Gaggioli, A., Pioggia, G., Tartarisco, G., Baldus, G., Corda, D., Cipresso, P. & Riva, G. (2013). A Mobile Data Collection Platform for Mental Health Research. *Personal and Ubiquitous Computing*, 17(2), 241–51. doi: 10.1007/s00779-011-0465-2.

[26]: Gomez Quiñonez, S., Walthouwer, M. J. L., Schulz, D. N. & de Vries, H. (2016). MHealth or EHealth? Efficacy, Use, and Appreciation of a Computer-Based Computer-Tailored Physical Activity Intervention for Dutch Adults: A Randomized Controlled Trial. *Journal of Medical Internet Research*, 18(11). doi: 10.2196/jmir.6171.

[27]: Gwaltney, C. J., Shields, A. L. & Shiffman, S. (2008). Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures: A Meta-Analytic Review. *Value in Health*, 11(2), 322–33. doi: 10.1111/j.1524-4733.2007.00231.x.

[28]: Haddadi, H., Hänsel, K. & Alomainy, A. (2016). Large scale mood and stress self-assessments on a smartwatch, *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16. ACM, New York, NY, USA, 1180–84. doi: 10.1145/2968219.2968305.

[29]: Healey, J. A., and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2), 156–66. doi: 10.1109/TITS.2005.848368.

[30]: Hedges, S. M., Jandorf, L. & Stone, A. A. (1985). Meaning of daily mood assessments. *Journal of Personality and Social Psychology,* 48(2), 428–34. doi: 10.1037/0022-3514.48.2.428.

[31]: Hektner, J. M., Schmidt, J. A. & Csikszentmihalyi, M. (2006). Experience Sampling Method: Measuring the Quality of Everyday Life. SAGE Publications Inc.

[32]: Hendriks, M., Ludwigs, K. & Veenhoven, R. (2016). Why Are Locals Happier than Internal Migrants? The Role of Daily Life. *Social Indicators Research,* 125(2), 481–508. doi: 10.1007/s11205-014-0856-7.

[33]: Hernandez, J., McDuff, D., Infante, C., Maes, P., Quigley, K. & Picard, R. (2016). Wearable ESM: Differences in the Experience Sampling Method Across Wearable Devices. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services,* MobileHCI '16. New York, NY, USA: ACM, 195–205. doi: 10.1145/2935334.2935340.

[34]: Hielckert, S. (2016). Measuring Happiness. Available at: http://hdl.handle.net/2105/34262 [Accessed: 21.07.2019].

[35]: Ice, G. and Gary, J. (2006). Measuring Stress in Humans. A Practical Guide for Field Research. Cambridge University Press.

[36]: Juster, T. F., and Stafford, F. P. (1985). *Time, goods, and well-being*. Ann Arbor, Mich.: Survey Research Center, Institute for Social Research, University of Michigan.

[37]: Kahneman, D., Krueger, A. B., Schkade, D. A. & Stone, A. A. (2004). The Day Reconstruction Method (DRM): Instrument Documentation. Available at: https://science.sciencemag.org/content/suppl/2004/12/01/306.5702.1776.DC1 [Accessed: 21.07.2019].

[38]: Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N. & Stone, A. A. (2004). A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science*, 306(5702), 1776–80. doi: 10.1126/science.1103572.

[39]: Kim, J., Kikuchi, H. & Yamamoto, Y. (2013). Systematic Comparison between Ecological Momentary Assessment and Day Reconstruction Method for Fatigue and Mood States in Healthy Adults. *British Journal of Health Psychology,* 18(1), 155–67. doi: 10.1111/bjhp.12000.

[40]: Krueger, A. B, and Schkade, D. A. (2008). The Reliability of Subjective Well-Being Measures. *Journal of Public Economics*, 92(8-9), 1833-1845. doi: 10.3386/w13027.

[41]: Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C. & Andreoli, A. (1993). Development of the perceived stress questionnaire: A new tool for psychosomatic research. *Journal of Psychosomatic Research,* 37(1), 19–32. doi: 10.1016/0022-3999(93)90120-5.

[42]: LLC, Yottaram (2019). Mood Tracker by EMoods - eMoods Insights. Available at: https://emoodtracker.com [Accessed: 21.07.2019].

[43]: MacLean, D., Roseway, A. & Czerwinski, M. (2013). MoodWings: A Wearable Biofeedback Device for Real-time Stress Intervention. *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments*, PETRA '13. New York, NY, USA: ACM, 66:1–66:8. doi: 10.1145/2504335.2504406.

[44]: McEwen, B. S., and Stellar, E. (1993). Stress and the Individual: Mechanisms Leading to Disease. *Archives of Internal Medicine,* 153(18), 2093–2101. doi: 10.1001/archinte.1993.00410180039004.

[45]: Morrison, L. G., Geraghty, A. W. A., Lloyd, S., Goodman, N., Michaelides, D. T., Hargood, C., Weal, M. & Yardley, L. (2018). Comparing usage of a computer and app stress management intervention: An observational study. *Internet Interventions,* 12, 74–82. doi: 10.1016/j.invent.2018.03.006.

[46]: Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P. & van Os, J. (2009). Experience Sampling Research in Psychopathology: Opening the Black Box of Daily Life. *Psychological Medicine,* 39(9), 1533–47. doi: 10.1017/S0033291708004947.

[47]: Myin-Germeys, I., Peeters, F., Havermans, R., Nicolson, N. A., DeVries, M. W., Delespaul, P. & Van Os, J. (2003). Emotional Reactivity to Daily Life Stress in Psychosis and Affective Disorder: An Experience Sampling Study. *Acta Psychiatrica Scandinavica,* 107(2), 124–31. doi: 10.1034/j.1600-0447.2003.02025.x.

[48]: Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W. & Reininghaus, U. (2018). Experience Sampling Methodology in Mental Health Research: New Insights and Technical Developments. *World Psychiatry,* 17(2), 123-132. doi: 10.1002/wps.20513.

[48]: Myin-Germeys, I., Klippel, A., Steinhart, H. & Reininghaus, U. (2016). Ecological Momentary Interventions in Psychiatry. *Current Opinion in Psychiatry*, 29(4), 258–63. doi: 10.1097/YCO.0000000000000255.

[50]: Norman, K. L. (2017). *Cyberpsychology: An Introduction to Human-Computer Interaction*. Cambridge University Press.

[51]: OneSignal (2019). High Volume Mobile and Computer Push Notifications. Available at: https://onesignal.com/ [Accessed: 21.07.2019].

[52]: Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P., Lewis, S. W. & Dunn, G. (2011). Experience Sampling Research in Individuals with Mental Illness: Reflections and Guidance. *Acta Psychiatrica Scandinavica*, 123(1), 12–20. doi: 10.1111/j.1600-0447.2010.01596.x.

[53]: Paredes, P. E., and Chan, M. (2011). CalmMeNow: Exploratory Research and Design of Stress Mitigating Mobile Interventions. *Proceedings of the International Conference on Human Factors in Computing Systems*, 1699–1704. doi: 10.1145/1979742.1979831.

[54]: Paredes, P., Gilad-Bachrach, R., Czerwinski, M., Roseway, A., Rowan, K. & Hernandez, J. (2014). PopTherapy: Coping with Stress Through Pop-culture. *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '14. ICST, Brussels, Belgium, Belgium: ICST, 109–117. doi: 10.4108/icst.pervasivehealth.2014.255070.

[55]: Pejovic, V., and Musolesi, M. (2015). Anticipatory Mobile Computing: A Survey of the State of the Art and Research Challenges. *ACM Computer Survey*, 47(3), 47:1–47:29. doi: 10.1145/2693843.

[56]: Plutchik, R. (2001). The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4), 344–50.

[57]: Randall, W. M., and Rickard, N. S. (2013). Development and Trial of a Mobile Experience Sampling Method (m-ESM) for Personal Music Listening. *Music Perception: An Interdisciplinary Journal*, 31(2), 157–70. doi: 10.1525/mp.2013.31.2.157.

[58]: Redelmeier, D. A., and Kahneman, D. (1996). Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures. *Pain*, 66(1), 3–8.

[59]: Reis, H., Gable, S. & Maniaci, M. (2014). Methods for studying everyday experience in its natural context. In Reis, H. & Judd, C. M. (Eds.), *Handbook of research methods in social and personality psychology* (pp. 373-403). New York, NY, US: Cambridge University Press.

[60]: Robinson, J. P., and Godbey, G. (1997). Time for Life: The Surprising Ways Americans Use Their Time. *American Journal of Sociology*, 104(1), 246–48. doi: 10.1086/210016.

[61]: Rubin, D. C., and Talarico, J. M. (2009). A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory*, 17(8), 802–8. doi: 10.1080/09658210903130764.

[62]: Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–78. doi: 10.1037/h0077714.

[63]: Sanvello Health (2019). Sanvello - On-demand help for stress, anxiety, and depression. Available at: https://www.sanvello.com/ [Accessed: 21.07.2019].

[64]: Self-Help for Anxiety Management (2015). Available at: https://apps.apple.com/us/app/self-help-for-anxiety-management/id666767947 [Accessed: 21.07.2019].

[65]: Shiffman, S. (2009). Ecological Momentary Assessment (EMA) in Studies of Substance Use. *Psychological assessment*, 21(4), 486–97. doi: 10.1037/a0017074.

[66]: Shiffman, S., Stone, A. A. & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4(1), 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415.

[67]: Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., Paty, J., Porter, L. S. & Cruise, L. J. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology*, 74(6), 1670–80. doi: 10.1037/0022-3514.74.6.1670.

[68]: Stone, A. A., and Shiffman, S. (1994). Ecological Momentary Assessment. *Annals of Behavioral Medicine*, 16(3), 199–202. doi: 10.1093/abm/16.3.199.

[69]: Stone, A., Shiffman, S., & DeVries, M. (1999). Ecological Momentary Assessment. In Kahneman D., Diener E., & Schwarz N. (Eds.), *Well-Being: Foundations of Hedonic Psychology* (pp. 26-39). Russell Sage Foundation. Available at: http://www.jstor.org/stable/10.7758/9781610443258.6 [Accessed: 21.07.2019].

[70]: Stoyanov, S. R, Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D. & Mani, M. (2015). Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. *JMIR mHealth and uHealth,* 3(1). doi: 10.2196/mhealth.3422.

[71]: T2 Mood Tracker. Available at: https://apps.apple.com/us/app/t2-mood-tracker/id428373825 [Accessed: 21.07.2019].

[72]: Tay, L., Chan, D. & Diener, E. (2014). The Metrics of Societal Happiness. *Social Indicators Research,* 117(2), 577–600. doi: 10.1007/s11205-013-0356-1.

[73]: Walls, T. and Schafer, J. L. (2006). Models for Intensive Longitudinal Data. Oxford University Press. doi: 10.1093/acprof:oso/9780195173444.001.0001.

[74]: Weppner, J., Lukowicz, P., Serino, S., Cipresso, P., Gaggioli, A., & Riva, G. (2013). Smartphone based experience sampling of stress-related events. *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 464–467. doi: 10.4108/icst.pervasivehealth.2013.252358.

[75]: Wilson, M., Doyle, J. & McTaggar, G. (2015). Designing Stress Management Interventions for Older Adults to Improve Wellbeing. *Proceedings of the 2015 British HCI Conference*, British HCI '15. New York, NY, USA: ACM, 307–308. doi: 10.1145/2783446.2783624.

[76]: World Health Organization (2019). Burn-out an occupational phenomenon: International Classification of Diseases. Available at: http://www.who.int/mental_health/evidence/burn-out/en/ [Accessed: 21.07.2019].

Appendix A: About your Day Application

Mobile Applications Review

Application Name		Reputation / number of times mentioned
Daylio		IIIII
T2 Mood Tracker eMoods Bipolar Mood Tracker		IIII
Pacifica / Sanvello Self-Help for Anxiety Management iMoodJournal Breathe2Relax aiMei Headspace Stigma Smiling Mind		III
RealifeChange MindShift Moods Happify Mindshift		II
M3 PTDS Coach Optimism Priori Therapy Buddy HelloMind Anxiety Reliever MoodTracking Diary What's up? Year in Pixels Stories – Timeline Diary Diary Mood Tracker Mood Log Timeline Mood Notes Up! MoodPanda	Mood Tracker Grid Diary Through Diary Pro Moodpath Mood & Anxiety Diary Moodcast Diary Moodtrack Social Diary Stop,Breathe & Think Calm 7 cups Digipill GPS for the Soul Stress Doctor Personal Zen MoodKit Worry Watch Mood Track Diary	I

Sources:

- https://www.everydayhealth.com/columns/therese-borchard-sanity-break/the-6-best-mood-apps/
- https://www.refinery29.com/en-us/best-mood-tracker-apps
- https://techzillo.com/best-mood-tracker-app/
- https://www.confidentlife.com.au/the-6-best-apps-to-track-your-mood/
- https://www.techuntold.com/best-mood-tracker-apps/
- https://thiswayup.org.au/12-free-apps-to-help-you-beat-stress/
- https://www.inc.com/lolly-daskal/13-of-the-best-apps-to-manage-stress.html
- https://www.healthline.com/health/anxiety/top-iphone-android-apps
- https://www.happierhuman.com/best-mood-tracker-apps/
- https://techwiser.com/best-mood-tracker-apps/

Mobile Application Rating Scale (MARS)

App Classification

The Classification section is used to collect descriptive and technical information about the app. Please review the app description in iTunes / Google Play to access this information.

App Name: Daylio

Rating this version: - **Rating all versions:** 4.6

Developer: Relaxio s.r.o

N ratings this version: **N ratings all versions:** 241

Version: 1.13.2 **Last update:** 10th July 2019

Cost - basic version: free **Cost - upgrade version:** -

Platform: ☐ iPhone ☐ iPad ☐ Android

Brief description: Daylio enables you to keep a private journal without having to type a single line.

Focus: what the app targets (select all that apply)

- ☐ Increase Happiness/Well-being
- ☐ Mindfulness/Meditation/Relaxation
- ☐ Reduce negative emotions
- ☐ Depression
- ☐ Anxiety/Stress
- ☐ Anger
- ☐ Behaviour Change
- ☐ Alcohol /Substance Use
- ☐ Goal Setting
- ☐ Entertainment
- ☐ Relationships
- ☐ Physical health
- ☐ Other

Theoretical background/Strategies (all that apply)

- ☐ Assessment
- ☐ Feedback
- ☐ Information/Education
- ☐ Monitoring/Tracking
- ☐ Goal setting
- ☐ Advice /Tips /Strategies /Skills training
- ☐ CBT - Behavioural (positive events)
- ☐ CBT – Cognitive (thought challenging)
- ☐ ACT - Acceptance commitment therapy
- ☐ Mindfulness/Meditation
- ☐ Relaxation
- ☐ Gratitude
- ☐ Strengths based
- ☐ Other

Affiliations:

☐ Unknown ☐ Commercial ☐ Government ☐ NGO ☐ University

Age group (all that apply)

- ☐ Children (under 12)
- ☐ Adolescents (13-17)
- ☐ Young Adults (18-25)
- ☐ Adults
- ☐ General

Technical aspects of app (all that apply)

- ☐ Allows sharing (Facebook, Twitter, etc.)
- ☐ Has an app community
- ☐ Allows password-protection
- ☐ Requires login
- ☐ Sends reminders
- ☐ Needs web access to function

App Quality Ratings

The Rating scale assesses app quality on four dimensions. All items are rated on a 5-point scale from "1.Inadequate" to "5.Excellent". Circle the number that most accurately represents the quality of the app component you are rating. Please use the descriptors provided for each response category.

SECTION A

Engagement – fun, interesting, customisable, interactive (e.g. sends alerts, messages, reminders, feedback, enables sharing), well-targeted to audience

1. **Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment (e.g. through gamification)?**
   - 1 Dull, not fun or entertaining at all
   - 2 Mostly boring
   - 3 OK, fun enough to entertain user for a brief time (< 5 minutes)
   - 4 Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
   - 5 Highly entertaining and fun, would stimulate repeat use

2. **Interest: Is the app interesting to use? Does it use any strategies to increase engagement by presenting its content in an interesting way?**
   - 1 Not interesting at all
   - 2 Mostly uninteresting
   - 3 OK, neither interesting nor uninteresting; would engage user for a brief time (< 5 minutes)
   - 4 Moderately interesting; would engage user for some time (5-10 minutes total)
   - 5 Very interesting, would engage user in repeat use

3. **Customisation: Does it provide/retain all necessary settings/preferences for apps features (e.g. sound, content, notifications, etc.)?**
   - 1 Does not allow any customisation or requires setting to be input every time
   - 2 Allows insufficient customisation limiting functions
   - 3 Allows basic customisation to function adequately
   - 4 Allows numerous options for customisation
   - 5 Allows complete tailoring to the individual's characteristics/preferences, retains all settings

4. **Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Note: these functions need to be customisable and not overwhelming in order to be perfect.**
   - 1 No interactive features and/or no response to user interaction
   - 2 Insufficient interactivity, or feedback, or user input options, limiting functions
   - 3 Basic interactive features to function adequately
   - 4 Offers a variety of interactive features/feedback/user input options
   - 5 Very high level of responsiveness through interactive features/feedback/user input options

5. **Target group: Is the app content (visual information, language, design) appropriate for your target audience?**
   - 1 Completely inappropriate/unclear/confusing
   - 2 Mostly inappropriate/unclear/confusing
   - 3 Acceptable but not targeted. May be inappropriate/unclear/confusing
   - 4 Well-targeted, with negligible issues
   - 5 Perfectly targeted, no issues found

A. Engagement mean score = 3.6

SECTION B

Functionality – app functioning, easy to learn, navigation, flow logic, and gestural design of app

6. **Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?**
   1 App is broken; no/insufficient/inaccurate response (e.g. crashes/bugs/broken features, etc.)
   2 Some functions work, but lagging or contains major technical problems
   3 App works overall. Some technical problems need fixing/Slow at times
   4 Mostly functional with minor/negligible problems
   5 Perfect/timely response, no technical bugs found/contains a 'loading time left' indicator

7. **Ease of use: How easy is it to learn how to use the app; how clear are the menu labels/icons and instructions?**
   1 No/limited instructions; menu labels/icons are confusing; complicated
   2 Useable after a lot of time/effort
   3 Useable after some time/effort
   4 Easy to learn how to use the app (or has clear instructions)
   5 Able to use app immediately; intuitive; simple

8. **Navigation: Is moving between screens logical/accurate/appropriate/ uninterrupted; are all necessary screen links present?**
   1 Different sections within the app seem logically disconnected and random/confusing/navigation is difficult
   2 Usable after a lot of time/effort
   3 Usable after some time/effort
   4 Easy to use or missing a negligible link
   5 Perfectly logical, easy, clear and intuitive screen flow throughout, or offers shortcuts

9. **Gestural design: Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?**
   1 Completely inconsistent/confusing
   2 Often inconsistent/confusing
   3 OK with some inconsistencies/confusing elements
   4 Mostly consistent/intuitive with negligible problems
   5 Perfectly consistent and intuitive

B. Functionality mean score = 5

SECTION C

Aesthetics – graphic design, overall visual appeal, colour scheme, and stylistic consistency

10. **Layout: Is arrangement and size of buttons/icons/menus/content on the screen appropriate or zoomable if needed?**
    1 Very bad design, cluttered, some options impossible to select/locate/see/read device display not optimised
    2 Bad design, random, unclear, some options difficult to select/locate/see/read
    3 Satisfactory, few problems with selecting/locating/seeing/reading items or with minor screen-size problems
    4 Mostly clear, able to select/locate/see/read items not zoomable
    5 Professional, simple, clear, orderly, logically organised, device display optimised. Every design component has a purpose

11. **Graphics: How high is the quality/resolution of graphics used for buttons/icons/menus/content?**
    1 Graphics appear amateur, very poor visual design - disproportionate, completely stylistically inconsistent
    2 Low quality/low resolution graphics; low quality visual design – disproportionate, stylistically inconsistent
    3 Moderate quality graphics and visual design (generally consistent in style)
    4 High quality/resolution graphics and visual design – mostly proportionate, stylistically consistent
    5 Very high quality/resolution graphics and visual design - proportionate, stylistically consistent throughout

12. **Visual appeal: How good does the app look?**
    1 No visual appeal, unpleasant to look at, poorly designed, clashing/mismatched colours
    2 Little visual appeal – poorly designed, bad use of colour, visually boring
    3 Some visual appeal – average, neither pleasant, nor unpleasant colors seem to be randomly chosen, too colourful
    4 High level of visual appeal – seamless graphics - consistent and professionally designed
    5 As above + very attractive, memorable, stands out; use of colour enhances app features/menus

C. Aesthetics mean score = 4.3

SECTION D

Information – Contains high quality information (e.g. text, feedback, measures, references) from a credible source. Select N/A if the app component is irrelevant.

13. **Accuracy of app description (in app store): Does app contain what is described?**
    1 Misleading. App does not contain the described components/functions. Or has no description
    2 Inaccurate. App contains very few of the described components/functions
    3 OK. App contains some of the described components/functions
    4 Accurate. App contains most of the described components/functions
    5 Highly accurate description of the app components/functions

14. **Goals: Does app have specific, measurable and achievable goals (specified in app store description or within the app itself)?**
    N/A Description does not list goals, or app goals are irrelevant to research goal (e.g. using a game for educational purposes)
    1 App has no chance of achieving its stated goals
    2 Description lists some goals, but app has very little chance of achieving them
    3 OK. App has clear goals, which may be achievable.
    4 App has clearly specified goals, which are measurable and achievable
    5 App has specific and measurable goals, which are highly likely to be achieved

15. **Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?**
    N/A There is no information within the app
    1 Irrelevant/inappropriate/incoherent/incorrect
    2 Poor. Barely relevant/appropriate/coherent/may be incorrect
    3 Moderately relevant/appropriate/coherent/and appears correct
    4 Relevant/appropriate/coherent/correct
    5 Highly relevant, appropriate, coherent, and correct

16. **Quantity of information: Is the extent coverage within the scope of the app; and comprehensive but concise?**

N/A There is no information within the app
1 Minimal or overwhelming
2 Insufficient or possibly overwhelming
3 OK but not comprehensive or concise
4 Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources
5 Comprehensive and concise; contains links to more information and resources

17. **Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc. – clear, logical, correct?**

N/A There is no visual information within the app (e.g. it only contains audio, or text)
1 Completely unclear/confusing/wrong or necessary but missing
2 Mostly unclear/confusing/wrong
3 OK but often unclear/confusing/wrong
4 Mostly clear/logical/correct with negligible issues
5 Perfectly clear/logical/correct

18. **Credibility: Does the app come from a legitimate source (specified in app store description or within the app itself)?**

1 Source identified but legitimacy/trustworthiness of source is questionable (e.g. commercial business with vested interest)
2 Appears to come from a legitimate source, but it cannot be verified (e.g. has no webpage)
3 Developed by small NGO/institution (hospital/centre, etc.) /specialised commercial business, funding body
4 Developed by government, university or as above but larger in scale
5 Developed using nationally competitive government or research funding (e.g. Australian Research Council, NHMRC)

19. **Evidence base: Has the app been trialled/tested; must be verified by evidence (in published scientific literature)?**

N/A The app has not been trialled/tested
1 The evidence suggests the app does not work
2 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has partially positive outcomes in studies that are not randomised controlled trials (RCTs), or there is little or no contradictory evidence.
3 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has positive outcomes in studies that are not RCTs, and there is no contradictory evidence.
4 App has been trialled and outcome tested in 1-2 RCTs indicating positive results
5 App has been trialled and outcome tested in ≥ 3 high quality RCTs indicating positive results

D. Information mean score = 3.7 *

* Exclude questions rated as "N/A" from the mean score calculation.

App subjective quality

SECTION E

20. **Would you recommend this app to people who might benefit from it?**

1 **Not at all** I would not recommend this app to anyone
2 There are very few people I would recommend this app to
3 **Maybe** There are several people whom I would recommend it to
4 There are many people I would recommend this app to
5 **Definitely** I would recommend this app to everyone

21. **How many times do you think you would use this app in the next 12 months if it was relevant to you?**

1 **None**
2 1-2
3 3-10
4 10-50
5 >50

22. **Would you pay for this app?**

1 No
3 Maybe
5 Yes

23. **What is your overall star rating of the app?**

1 ★ One of the worst apps I've used
2 ★★
3 ★★★ Average
4 ★★★★
5 ★★★★★ One of the best apps I've used

Scoring

App quality scores for

SECTION

A: Engagement Mean Score = 3.6
B: Functionality Mean Score = 5
C: Aesthetics Mean Score = 4.3
D: Information Mean Score = 3.7
App quality mean Score = 4,1
App subjective quality Score = 3

Mobile Application Rating Scale (MARS)

App Classification

The Classification section is used to collect descriptive and technical information about the app. Please review the app description in iTunes / Google Play to access this information.

App Name: eMoods Bipolar Mood Tracker

Rating this version: - **Rating all versions:** 4.8

Developer: Yottaram,LLC

N ratings this version: - **N ratings all versions:** 10

Version: 1.2.9 **Last update:** 10.07.2019

Cost - basic version: free **Cost - upgrade version:** -

Platform: ☐ iPhone ☐ iPad ☐ Android

Brief description: Track your mood and other symptoms in seconds effortlessly

Focus: what the app targets (select all that apply)

- ☐ Increase Happiness/Well-being
- ☐ Mindfulness/Meditation/Relaxation
- ☐ Reduce negative emotions
- ☐ Depression
- ☐ Anxiety/Stress
- ☐ Anger
- ☐ Behaviour Change
- ☐ Alcohol /Substance Use
- ☐ Goal Setting
- ☐ Entertainment
- ☐ Relationships
- ☐ Physical health
- ☐ Other _____

Theoretical background/Strategies (all that apply)

- ☐ Assessment
- ☐ Feedback
- ☐ Information/Education
- ☐ Monitoring/Tracking
- ☐ Goal setting
- ☐ Advice /Tips /Strategies /Skills training
- ☐ CBT - Behavioural (positive events)
- ☐ CBT – Cognitive (thought challenging)
- ☐ ACT - Acceptance commitment therapy
- ☐ Mindfulness/Meditation
- ☐ Relaxation
- ☐ Gratitude
- ☐ Strengths based
- ☐ Other _____

Affiliations:

☐ Unknown ☐ Commercial ☐ Government ☐ NGO ☐ University

Age group (all that apply)

- ☐ Children (under 12)
- ☐ Adolescents (13-17)
- ☐ Young Adults (18-25)
- ☐ Adults
- ☐ General

Technical aspects of app (all that apply)

- ☐ Allows sharing (Facebook, Twitter, etc.)
- ☐ Has an app community
- ☐ Allows password-protection
- ☐ Requires login
- ☐ Sends reminders
- ☐ Needs web access to function

App Quality Ratings

The Rating scale assesses app quality on four dimensions. All items are rated on a 5-point scale from "1.Inadequate" to "5.Excellent". Circle the number that most accurately represents the quality of the app component you are rating. Please use the descriptors provided for each response category.

SECTION A

Engagement – fun, interesting, customisable, interactive (e.g. sends alerts, messages, reminders, feedback, enables sharing), well-targeted to audience

1. **Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment (e.g. through gamification)?**
   - 1 Dull, not fun or entertaining at all
   - 2 Mostly boring
   - 3 OK, fun enough to entertain user for a brief time (< 5 minutes)
   - 4 Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
   - 5 Highly entertaining and fun, would stimulate repeat use

2. **Interest: Is the app interesting to use? Does it use any strategies to increase engagement by presenting its content in an interesting way?**
   - 1 Not interesting at all
   - 2 Mostly uninteresting
   - 3 OK, neither interesting nor uninteresting; would engage user for a brief time (< 5 minutes)
   - 4 Moderately interesting; would engage user for some time (5-10 minutes total)
   - 5 Very interesting, would engage user in repeat use

3. **Customisation: Does it provide/retain all necessary settings/preferences for apps features (e.g. sound, content, notifications, etc.)?**
   - 1 Does not allow any customisation or requires setting to be input every time
   - 2 Allows insufficient customisation limiting functions
   - 3 Allows basic customisation to function adequately
   - 4 Allows numerous options for customisation
   - 5 Allows complete tailoring to the individual's characteristics/preferences, retains all settings

4. **Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Note: these functions need to be customisable and not overwhelming in order to be perfect.**
   - 1 No interactive features and/or no response to user interaction
   - 2 Insufficient interactivity, or feedback, or user input options, limiting functions
   - 3 Basic interactive features to function adequately
   - 4 Offers a variety of interactive features/feedback/user input options
   - 5 Very high level of responsiveness through interactive features/feedback/user input options

5. **Target group: Is the app content (visual information, language, design) appropriate for your target audience?**
   - 1 Completely inappropriate/unclear/confusing
   - 2 Mostly inappropriate/unclear/confusing
   - 3 Acceptable but not targeted. May be inappropriate/unclear/confusing
   - 4 Well-targeted, with negligible issues
   - 5 Perfectly targeted, no issues found

A. Engagement mean score = 3

SECTION B

Functionality – app functioning, easy to learn, navigation, flow logic, and gestural design of app

6. **Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?**
   1 App is broken; no/insufficient/inaccurate response (e.g. crashes/bugs/broken features, etc.)
   2 Some functions work, but lagging or contains major technical problems
   3 App works overall. Some technical problems need fixing/Slow at times
   4 Mostly functional with minor/negligible problems
   5 Perfect/timely response; no technical bugs found/contains a 'loading time left' indicator

7. **Ease of use: How easy is it to learn how to use the app; how clear are the menu labels/icons and instructions?**
   1 No/limited instructions; menu labels/icons are confusing; complicated
   2 Useable after a lot of time/effort
   3 Useable after some time/effort
   4 Easy to learn how to use the app (or has clear instructions)
   5 Able to use app immediately; intuitive; simple

8. **Navigation: Is moving between screens logical/accurate/appropriate/ uninterrupted; are all necessary screen links present?**
   1 Different sections within the app seem logically disconnected and random/confusing/navigation is difficult
   2 Usable after a lot of time/effort
   3 Usable after some time/effort
   4 Easy to use or missing a negligible link
   5 Perfectly logical, easy, clear and intuitive screen flow throughout, or offers shortcuts

9. **Gestural design: Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?**
   1 Completely inconsistent/confusing
   2 Often inconsistent/confusing
   3 OK with some inconsistencies/confusing elements
   4 Mostly consistent/intuitive with negligible problems
   5 Perfectly consistent and intuitive

B. Functionality mean score = 4.25

SECTION C

Aesthetics – graphic design, overall visual appeal, colour scheme, and stylistic consistency

10. **Layout: Is arrangement and size of buttons/icons/menus/content on the screen appropriate or zoomable if needed?**
    1 Very bad design, cluttered, some options impossible to select/locate/see/read device display not optimised
    2 Bad design, random, unclear, some options difficult to select/locate/see/read
    3 Satisfactory, few problems with selecting/locating/seeing/reading items or with minor screen-size problems
    4 Mostly clear, able to select/locate/see/read items font too small
    5 Professional, simple, clear, orderly, logically organised, device display optimised. Every design component has a purpose

11. **Graphics: How high is the quality/resolution of graphics used for buttons/icons/menus/content?**
    1 Graphics appear amateur, very poor visual design - disproportionate, completely stylistically inconsistent
    2 Low quality/low resolution graphics; low quality visual design – disproportionate, stylistically inconsistent
    3 Moderate quality graphics and visual design (generally consistent in style)
    4 High quality/resolution graphics and visual design – mostly proportionate, stylistically consistent
    5 Very high quality/resolution graphics and visual design - proportionate, stylistically consistent throughout

12. **Visual appeal: How good does the app look?**
    1 No visual appeal, unpleasant to look at, poorly designed, clashing/mismatched colours
    2 Little visual appeal – poorly designed, bad use of colour, visually boring
    3 Some visual appeal – average, neither pleasant, nor unpleasant
    4 High level of visual appeal – seamless graphics – consistent and professionally designed
    5 As above + very attractive, memorable, stands out; use of colour enhances app features/menus

C. Aesthetics mean score = 4.3

SECTION D

Information – Contains high quality information (e.g. text, feedback, measures, references) from a credible source. Select N/A if the app component is irrelevant.

13. **Accuracy of app description (in app store): Does app contain what is described?**
    1 Misleading. App does not contain the described components/functions. Or has no description
    2 Inaccurate. App contains very few of the described components/functions
    3 OK. App contains some of the described components/functions
    4 Accurate. App contains most of the described components/functions
    5 Highly accurate description of the app components/functions

14. **Goals: Does app have specific, measurable and achievable goals (specified in app store description or within the app itself)?**
    N/A Description does not list goals, or app goals are irrelevant to research goal (e.g. using a game for educational purposes)
    1 App has no chance of achieving its stated goals
    2 Description lists some goals, but app has very little chance of achieving them
    3 OK. App has clear goals, which may be achievable.
    4 App has clearly specified goals, which are measurable and achievable
    5 App has specific and measurable goals, which are highly likely to be achieved

15. **Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?**
    N/A There is no information within the app
    1 Irrelevant/inappropriate/incoherent/incorrect
    2 Poor. Barely relevant/appropriate/coherent/may be incorrect
    3 Moderately relevant/appropriate/coherent/and appears correct
    4 Relevant/appropriate/coherent/correct
    5 Highly relevant, appropriate, coherent, and correct

16. **Quantity of information: Is the extent coverage within the scope of the app; and comprehensive but concise?**

N/A There is no information within the app
1 Minimal or overwhelming
2 Insufficient or possibly overwhelming
3 OK but not comprehensive or concise
4 Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources
5 Comprehensive and concise; contains links to more information and resources

17. **Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc. – clear, logical, correct?**

N/A There is no visual information within the app (e.g. it only contains audio, or text)
1 Completely unclear/confusing/wrong or necessary but missing
2 Mostly unclear/confusing/wrong
3 OK but often unclear/confusing/wrong
4 Mostly clear/logical/correct with negligible issues
5 Perfectly clear/logical/correct

18. **Credibility: Does the app come from a legitimate source (specified in app store description or within the app itself)?**

1 Source identified but legitimacy/trustworthiness of source is questionable (e.g. commercial business with vested interest)
2 Appears to come from a legitimate source, but it cannot be verified (e.g. has no webpage)
3 Developed by small NGO/institution (hospital/centre, etc.) /specialised commercial business, funding body
4 Developed by government, university or as above but larger in scale
5 Developed using nationally competitive government or research funding (e.g. Australian Research Council, NHMRC)

19. **Evidence base: Has the app been trialled/tested; must be verified by evidence (in published scientific literature)?**

N/A The app has not been trialled/tested
1 The evidence suggests the app does not work
2 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has partially positive outcomes in studies that are not randomised controlled trials (RCTs), or there is little or no contradictory evidence.
3 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has positive outcomes in studies that are not RCTs, and there is no contradictory evidence.
4 App has been trialled and outcome tested in 1-2 RCTs indicating positive results
5 App has been trialled and outcome tested in ≥ 3 high quality RCTs indicating positive results

D. Information mean score = 3.7 *

* Exclude questions rated as "N/A" from the mean score calculation.

App subjective quality

SECTION E

20. **Would you recommend this app to people who might benefit from it?**

1 **Not at all** I would not recommend this app to anyone
2 There are very few people I would recommend this app to
3 **Maybe** There are several people whom I would recommend it to
4 There are many people I would recommend this app to
5 **Definitely** I would recommend this app to everyone

21. **How many times do you think you would use this app in the next 12 months if it was relevant to you?**

1 **None**
2 1-2
3 3-10
4 10-50
5 >50

22. **Would you pay for this app?**

1 No
3 Maybe
5 Yes

23. **What is your overall star rating of the app?**

1 ★ One of the worst apps I've used
2 ★★
3 ★★★ Average
4 ★★★★
5 ★★★★★ One of the best apps I've used

Scoring

App quality scores for

SECTION

A: Engagement Mean Score = 3

B: Functionality Mean Score = 4.25

C: Aesthetics Mean Score = 4.3

D: Information Mean Score = 3.7

App quality mean Score = 3.8

App subjective quality Score = 1.8

Mobile Application Rating Scale (MARS)

App Classification

The Classification section is used to collect descriptive and technical information about the app. Please review the app description in iTunes / Google Play to access this information.

App Name: Sanvello (formerly Pacifica)

Rating this version: _____ **Rating all versions:** 3.5

Developer: Pacifica Labs Inc.

N ratings this version: _____ **N ratings all versions:** 13

Version: 8.0.2 **Last update:** 12.07.2019

Cost - basic version: free **Cost - upgrade version:** -

Platform: ☐ iPhone ☐ iPad ☐ Android

Brief description: Sanvello is on-demand help for stress, anxiety and depression. It offers clinically validated techniques and support to help you relieve symptoms and feel happier over time.

Focus: what the app targets (select all that apply)

- ☐ Increase Happiness/Well-being
- ☐ Mindfulness/Meditation/Relaxation
- ☐ Reduce negative emotions
- ☐ Depression
- ☐ Anxiety/Stress
- ☐ Anger
- ☐ Behaviour Change
- ☐ Alcohol /Substance Use
- ☐ Goal Setting
- ☐ Entertainment
- ☐ Relationships
- ☐ Physical health
- ☐ Other _____

Theoretical background/Strategies (all that apply)

- ☐ Assessment
- ☐ Feedback
- ☐ Information/Education
- ☐ Monitoring/Tracking
- ☐ Goal setting
- ☐ Advice /Tips /Strategies /Skills training
- ☐ CBT - Behavioural (positive events)
- ☐ CBT – Cognitive (thought challenging)
- ☐ ACT - Acceptance commitment therapy
- ☐ Mindfulness/Meditation
- ☐ Relaxation
- ☐ Gratitude
- ☐ Strengths based
- ☐ Other _____

Affiliations:

☐ Unknown ☐ Commercial ☐ Government ☐ NGO ☐ University

Age group (all that apply)

- ☐ Children (under 12)
- ☐ Adolescents (13-17)
- ☐ Young Adults (18-25)
- ☐ Adults
- ☐ General

Technical aspects of app (all that apply)

- ☐ Allows sharing (Facebook, Twitter, etc.)
- ☐ Has an app community
- ☐ Allows password-protection
- ☐ Requires login
- ☐ Sends reminders
- ☐ Needs web access to function

App Quality Ratings

The Rating scale assesses app quality on four dimensions. All items are rated on a 5-point scale from "1.Inadequate" to "5.Excellent". Circle the number that most accurately represents the quality of the app component you are rating. Please use the descriptors provided for each response category.

SECTION A

Engagement – fun, interesting, customisable, interactive (e.g. sends alerts, messages, reminders, feedback, enables sharing), well-targeted to audience

1. **Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment (e.g. through gamification)?**
   - 1 Dull, not fun or entertaining at all
   - 2 Mostly boring
   - 3 OK, fun enough to entertain user for a brief time (< 5 minutes)
   - 4 Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
   - 5 Highly entertaining and fun, would stimulate repeat use

2. **Interest: Is the app interesting to use? Does it use any strategies to increase engagement by presenting its content in an interesting way?**
   - 1 Not interesting at all
   - 2 Mostly uninteresting
   - 3 OK, neither interesting nor uninteresting; would engage user for a brief time (< 5 minutes)
   - 4 Moderately interesting; would engage user for some time (5-10 minutes total)
   - 5 Very interesting, would engage user in repeat use

3. **Customisation: Does it provide/retain all necessary settings/preferences for apps features (e.g. sound, content, notifications, etc.)?**
   - 1 Does not allow any customisation or requires setting to be input every time
   - 2 Allows insufficient customisation limiting functions
   - 3 Allows basic customisation to function adequately
   - 4 Allows numerous options for customisation
   - 5 Allows complete tailoring to the individual's characteristics/preferences, retains all settings

4. **Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Note: these functions need to be customisable and not overwhelming in order to be perfect.**
   - 1 No interactive features and/or no response to user interaction
   - 2 Insufficient interactivity, or feedback, or user input options, limiting functions
   - 3 Basic interactive features to function adequately
   - 4 Offers a variety of interactive features/feedback/user input options
   - 5 Very high level of responsiveness through interactive features/feedback/user input options

5. **Target group: Is the app content (visual information, language, design) appropriate for your target audience?**
   - 1 Completely inappropriate/unclear/confusing
   - 2 Mostly inappropriate/unclear/confusing
   - 3 Acceptable but not targeted. May be inappropriate/unclear/confusing
   - 4 Well-targeted, with negligible issues
   - 5 Perfectly targeted, no issues found

A. Engagement mean score = 4.8

SECTION B

Functionality – app functioning, easy to learn, navigation, flow logic, and gestural design of app

6. **Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?**
   1 App is broken; no/insufficient/inaccurate response (e.g. crashes/bugs/broken features, etc.)
   2 Some functions work, but lagging or contains major technical problems
   3 App works overall. Some technical problems need fixing/Slow at times
   4 Mostly functional with minor/negligible problems
   5 Perfect/timely response; no technical bugs found/contains a 'loading time left' indicator

7. **Ease of use: How easy is it to learn how to use the app; how clear are the menu labels/icons and instructions?**
   1 No/limited instructions; menu labels/icons are confusing; complicated
   2 Useable after a lot of time/effort
   3 Useable after some time/effort
   4 Easy to learn how to use the app (or has clear instructions) a little overloaded
   5 Able to use app immediately; intuitive; simple

8. **Navigation: Is moving between screens logical/accurate/appropriate/ uninterrupted; are all necessary screen links present?**
   1 Different sections within the app seem logically disconnected and random/confusing/navigation is difficult
   2 Usable after a lot of time/effort
   3 Usable after some time/effort
   4 Easy to use or missing a negligible link
   5 Perfectly logical, easy, clear and intuitive screen flow throughout, or offers shortcuts

9. **Gestural design: Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?**
   1 Completely inconsistent/confusing
   2 Often inconsistent/confusing
   3 OK with some inconsistencies/confusing elements
   4 Mostly consistent/intuitive with negligible problems
   5 Perfectly consistent and intuitive

B. Functionality mean score = 4.5

SECTION C

Aesthetics – graphic design, overall visual appeal, colour scheme, and stylistic consistency

10. **Layout: Is arrangement and size of buttons/icons/menus/content on the screen appropriate or zoomable if needed?**
    1 Very bad design, cluttered, some options impossible to select/locate/see/read device display not optimised
    2 Bad design, random, unclear, some options difficult to select/locate/see/read
    3 Satisfactory, few problems with selecting/locating/seeing/reading items or with minor screen-size problems
    4 Mostly clear, able to select/locate/see/read items overloaded
    5 Professional, simple, clear, orderly, logically organised, device display optimised. Every design component has a purpose

11. **Graphics: How high is the quality/resolution of graphics used for buttons/icons/menus/content?**
    1 Graphics appear amateur, very poor visual design - disproportionate, completely stylistically inconsistent
    2 Low quality/low resolution graphics; low quality visual design – disproportionate, stylistically inconsistent
    3 Moderate quality graphics and visual design (generally consistent in style)
    4 High quality/resolution graphics and visual design – mostly proportionate, stylistically consistent
    5 Very high quality/resolution graphics and visual design - proportionate, stylistically consistent throughout

12. **Visual appeal: How good does the app look?**
    1 No visual appeal, unpleasant to look at, poorly designed, clashing/mismatched colours
    2 Little visual appeal – poorly designed, bad use of colour, visually boring
    3 Some visual appeal – average, neither pleasant, nor unpleasant
    4 High level of visual appeal – seamless graphics – consistent and professionally designed
    5 As above + very attractive, memorable, stands out; use of colour enhances app features/menus

C. Aesthetics mean score = 4.3

SECTION D

Information – Contains high quality information (e.g. text, feedback, measures, references) from a credible source. Select N/A if the app component is irrelevant.

13. **Accuracy of app description (in app store): Does app contain what is described?**
    1 Misleading. App does not contain the described components/functions. Or has no description
    2 Inaccurate. App contains very few of the described components/functions
    3 OK. App contains some of the described components/functions
    4 Accurate. App contains most of the described components/functions
    5 Highly accurate description of the app components/functions

14. **Goals: Does app have specific, measurable and achievable goals (specified in app store description or within the app itself)?**
    N/A Description does not list goals, or app goals are irrelevant to research goal (e.g. using a game for educational purposes)
    1 App has no chance of achieving its stated goals
    2 Description lists some goals, but app has very little chance of achieving them
    3 OK. App has clear goals, which may be achievable.
    4 App has clearly specified goals, which are measurable and achievable
    5 App has specific and measurable goals, which are highly likely to be achieved

15. **Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?**
    N/A There is no information within the app
    1 Irrelevant/inappropriate/incoherent/incorrect
    2 Poor. Barely relevant/appropriate/coherent/may be incorrect
    3 Moderately relevant/appropriate/coherent/and appears correct
    4 Relevant/appropriate/coherent/correct
    5 Highly relevant, appropriate, coherent, and correct

16. **Quantity of information: Is the extent coverage within the scope of the app; and comprehensive but concise?**

N/A There is no information within the app
1 Minimal or overwhelming
2 Insufficient or possibly overwhelming
3 OK but not comprehensive or concise
4 Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources
5 Comprehensive and concise; contains links to more information and resources

17. **Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc. – clear, logical, correct?**

N/A There is no visual information within the app (e.g. it only contains audio, or text)
1 Completely unclear/confusing/wrong or necessary but missing
2 Mostly unclear/confusing/wrong
3 OK but often unclear/confusing/wrong
4 Mostly clear/logical/correct with negligible issues
5 Perfectly clear/logical/correct

18. **Credibility: Does the app come from a legitimate source (specified in app store description or within the app itself)?**

1 Source identified but legitimacy/trustworthiness of source is questionable (e.g. commercial business with vested interest)
2 Appears to come from a legitimate source, but it cannot be verified (e.g. has no webpage)
3 Developed by small NGO/institution (hospital/centre, etc.) /specialised commercial business, funding body
4 Developed by government, university or as above but larger in scale
5 Developed using nationally competitive government or research funding (e.g. Australian Research Council, NHMRC)

19. **Evidence base: Has the app been trialled/tested; must be verified by evidence (in published scientific literature)?**

N/A The app has not been trialled/tested
1 The evidence suggests the app does not work
2 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has partially positive outcomes in studies that are not randomised controlled trials (RCTs), or there is little or no contradictory evidence.
3 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has positive outcomes in studies that are not RCTs, and there is no contradictory evidence.
4 App has been trialled and outcome tested in 1-2 RCTs indicating positive results
5 App has been trialled and outcome tested in ≥ 3 high quality RCTs indicating positive results

D. Information mean score = 4.0 *

* Exclude questions rated as "N/A" from the mean score calculation.

App subjective quality

SECTION E

20. **Would you recommend this app to people who might benefit from it?**

1 **Not at all** I would not recommend this app to anyone
2 There are very few people I would recommend this app to
3 **Maybe** There are several people whom I would recommend it to
4 There are many people I would recommend this app to
5 **Definitely** I would recommend this app to everyone

21. **How many times do you think you would use this app in the next 12 months if it was relevant to you?**

1 **None**
2 1-2
3 3-10
4 10-50
5 >50

22. **Would you pay for this app?**

1 No
3 Maybe
5 Yes

23. **What is your overall star rating of the app?**

1 ★ One of the worst apps I've used
2 ★★
3 ★★★ Average
4 ★★★★
5 ★★★★★ One of the best apps I've used

Scoring

App quality scores for

SECTION

A: Engagement Mean Score = 4.8
B: Functionality Mean Score = 4.5
C: Aesthetics Mean Score = 4.3
D: Information Mean Score = 4.0
App quality mean Score = 4.4
App subjective quality Score = 4

Mobile Application Rating Scale (MARS)

App Classification

The Classification section is used to collect descriptive and technical information about the app. Please review the app description in iTunes / Google Play to access this information.

App Name: Self Anxiety Management

Rating this version: not enough ratings **Rating all versions:** not enough ratings

Developer: University of the West of England

N ratings this version: - **N ratings all versions:** -

Version: 1.2.8 **Last update:** 2017

Cost - basic version: free **Cost - upgrade version:** -

Platform: ☐ iPhone ☐ iPad ☐ Android

Brief description: SAM is a friendly app that offers a range of self-help methods for people who are serious about learning to manage their anxiety.

Focus: what the app targets (select all that apply)

- ☐ Increase Happiness/Well-being
- ☐ Mindfulness/Meditation/Relaxation
- ☐ Reduce negative emotions
- ☐ Depression
- ☐ Anxiety/Stress
- ☐ Anger
- ☐ Behaviour Change
- ☐ Alcohol /Substance Use
- ☐ Goal Setting
- ☐ Entertainment
- ☐ Relationships
- ☐ Physical health
- ☐ Other _____

Theoretical background/Strategies (all that apply)

- ☐ Assessment
- ☐ Feedback
- ☐ Information/Education
- ☐ Monitoring/Tracking
- ☐ Goal setting
- ☐ Advice /Tips /Strategies /Skills training
- ☐ CBT - Behavioural (positive events)
- ☐ CBT – Cognitive (thought challenging)
- ☐ ACT - Acceptance commitment therapy
- ☐ Mindfulness/Meditation
- ☐ Relaxation
- ☐ Gratitude
- ☐ Strengths based
- ☐ Other _____

Affiliations:

☐ Unknown ☐ Commercial ☐ Government ☐ NGO ☐ University

Age group (all that apply)

- ☐ Children (under 12)
- ☐ Adolescents (13-17)
- ☐ Young Adults (18-25)
- ☐ Adults
- ☐ General

Technical aspects of app (all that apply)

- ☐ Allows sharing (Facebook, Twitter, etc.)
- ☐ Has an app community
- ☐ Allows password protection
- ☐ Requires login
- ☐ Sends reminders
- ☐ Needs web access to function

App Quality Ratings

The Rating scale assesses app quality on four dimensions. All items are rated on a 5-point scale from "1.Inadequate" to "5.Excellent". Circle the number that most accurately represents the quality of the app component you are rating. Please use the descriptors provided for each response category.

SECTION A

Engagement – fun, interesting, customisable, interactive (e.g. sends alerts, messages, reminders, feedback, enables sharing), well-targeted to audience

1. **Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment (e.g. through gamification)?**
   1 Dull, not fun or entertaining at all
   2 Mostly boring
   3 OK, fun enough to entertain user for a brief time (< 5 minutes)
   4 Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
   5 Highly entertaining and fun, would stimulate repeat use

2. **Interest: Is the app interesting to use? Does it use any strategies to increase engagement by presenting its content in an interesting way?**
   1 Not interesting at all
   2 Mostly uninteresting
   3 OK, neither interesting nor uninteresting; would engage user for a brief time (< 5 minutes)
   4 Moderately interesting; would engage user for some time (5-10 minutes total)
   5 Very interesting, would engage user in repeat use

3. **Customisation: Does it provide/retain all necessary settings/preferences for apps features (e.g. sound, content, notifications, etc.)?**
   1 Does not allow any customisation or requires setting to be input every time
   2 Allows insufficient customisation limiting functions
   3 Allows basic customisation to function adequately
   4 Allows numerous options for customisation
   5 Allows complete tailoring to the individual's characteristics/preferences, retains all settings

4. **Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Note: these functions need to be customisable and not overwhelming in order to be perfect.**
   1 No interactive features and/or no response to user interaction
   2 Insufficient interactivity, or feedback, or user input options, limiting functions
   3 Basic interactive features to function adequately
   4 Offers a variety of interactive features/feedback/user input options
   5 Very high level of responsiveness through interactive features/feedback/user input options

5. **Target group: Is the app content (visual information, language, design) appropriate for your target audience?**
   1 Completely inappropriate/unclear/confusing
   2 Mostly inappropriate/unclear/confusing
   3 Acceptable but not targeted. May be inappropriate/unclear/confusing
   4 Well-targeted, with negligible issues
   5 Perfectly targeted, no issues found

A. Engagement mean score = 2.8

SECTION B

Functionality – app functioning, easy to learn, navigation, flow logic, and gestural design of app

6. **Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?**
   1 App is broken; no/insufficient/inaccurate response (e.g. crashes/bugs/broken features, etc.)
   2 Some functions work, but lagging or contains major technical problems
   3 App works overall. Some technical problems need fixing/Slow at times
   4 Mostly functional with minor/negligible problems
   5 Perfect/timely response; no technical bugs found/contains a 'loading time left' indicator

7. **Ease of use: How easy is it to learn how to use the app; how clear are the menu labels/icons and instructions?**
   1 No/limited instructions; menu labels/icons are confusing; complicated
   2 Useable after a lot of time/effort
   3 Useable after some time/effort
   4 Easy to learn how to use the app (or has clear instructions)
   5 Able to use app immediately; intuitive; simple

8. **Navigation: Is moving between screens logical/accurate/appropriate/ uninterrupted; are all necessary screen links present?**
   1 Different sections within the app seem logically disconnected and random/confusing/navigation is difficult
   2 Usable after a lot of time/effort
   3 Usable after some time/effort
   4 Easy to use or missing a negligible link
   5 Perfectly logical, easy, clear and intuitive screen flow throughout, or offers shortcuts

9. **Gestural design: Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?**
   1 Completely inconsistent/confusing
   2 Often inconsistent/confusing
   3 OK with some inconsistencies/confusing elements
   4 Mostly consistent/intuitive with negligible problems
   5 Perfectly consistent and intuitive

B. Functionality mean score = 4.25

SECTION C

Aesthetics – graphic design, overall visual appeal, colour scheme, and stylistic consistency

10. **Layout: Is arrangement and size of buttons/icons/menus/content on the screen appropriate or zoomable if needed?**
    1 Very bad design, cluttered, some options impossible to select/locate/see/read device display not optimised
    2 Bad design, random, unclear, some options difficult to select/locate/see/read
    3 Satisfactory, few problems with selecting/locating/seeing/reading items or with minor screen-size problems
    4 Mostly clear, able to select/locate/see/read items
    5 Professional, simple, clear, orderly, logically organised, device display optimised. Every design component has a purpose

11. **Graphics: How high is the quality/resolution of graphics used for buttons/icons/menus/content?**
    1 Graphics appear amateur, very poor visual design - disproportionate, completely stylistically inconsistent
    2 Low quality/low resolution graphics; low quality visual design – disproportionate, stylistically inconsistent
    3 Moderate quality graphics and visual design (generally consistent in style)
    4 High quality/resolution graphics and visual design – mostly proportionate, stylistically consistent
    5 Very high quality/resolution graphics and visual design - proportionate, stylistically consistent throughout

12. **Visual appeal: How good does the app look?**
    1 No visual appeal, unpleasant to look at, poorly designed, clashing/mismatched colours
    2 Little visual appeal – poorly designed, bad use of colour, visually boring
    3 Some visual appeal - average, neither pleasant nor unpleasant
    4 High level of visual appeal – seamless graphics – consistent and professionally designed
    5 As above + very attractive, memorable, stands out; use of colour enhances app features/menus

C. Aesthetics mean score = 4

SECTION D

Information – Contains high quality information (e.g. text, feedback, measures, references) from a credible source. Select N/A if the app component is irrelevant.

13. **Accuracy of app description (in app store): Does app contain what is described?**
    1 Misleading. App does not contain the described components/functions. Or has no description
    2 Inaccurate. App contains very few of the described components/functions
    3 OK. App contains some of the described components/functions
    4 Accurate. App contains most of the described components/functions
    5 Highly accurate description of the app components/functions

14. **Goals: Does app have specific, measurable and achievable goals (specified in app store description or within the app itself)?**
    N/A Description does not list goals, or app goals are irrelevant to research goal (e.g. using a game for educational purposes)
    1 App has no chance of achieving its stated goals
    2 Description lists some goals, but app has very little chance of achieving them
    3 OK. App has clear goals, which may be achievable.
    4 App has clearly specified goals, which are measurable and achievable
    5 App has specific and measurable goals, which are highly likely to be achieved

15. **Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?**
    N/A There is no information within the app
    1 Irrelevant/inappropriate/incoherent/incorrect
    2 Poor. Barely relevant/appropriate/coherent/may be incorrect
    3 Moderately relevant/appropriate/coherent/and appears correct
    4 Relevant/appropriate/coherent/correct
    5 Highly relevant, appropriate, coherent, and correct

16. **Quantity of information: Is the extent coverage within the scope of the app; and comprehensive but concise?**

N/A There is no information within the app
1 Minimal or overwhelming
2 Insufficient or possibly overwhelming
3 OK but not comprehensive or concise
4 Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources
5 Comprehensive and concise; contains links to more information and resources

17. **Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc – clear, logical, correct?**

N/A There is no visual information within the app (e.g. it only contains audio, or text)
1 Completely unclear/confusing/wrong or necessary but missing
2 Mostly unclear/confusing/wrong
3 OK but often unclear/confusing/wrong
4 Mostly clear/logical/correct with negligible issues
5 Perfectly clear/logical/correct

18. **Credibility: Does the app come from a legitimate source (specified in app store description or within the app itself)?**

1 Source identified but legitimacy/trustworthiness of source is questionable (e.g. commercial business with vested interest)
2 Appears to come from a legitimate source, but it cannot be verified (e.g. has no webpage)
3 Developed by small NGO/institution (hospital/centre, etc.) /specialised commercial business, funding body
4 Developed by government, university or as above but larger in scale
5 Developed using nationally competitive government or research funding (e.g. Australian Research Council, NHMRC)

19. **Evidence base: Has the app been trialled/tested; must be verified by evidence (in published scientific literature)?**

N/A The app has not been trialled/tested
1 The evidence suggests the app does not work
2 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has partially positive outcomes in studies that are not randomised controlled trials (RCTs), or there is little or no contradictory evidence.
3 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has positive outcomes in studies that are not RCTs, and there is no contradictory evidence.
4 App has been trialled and outcome tested in 1-2 RCTs indicating positive results
5 App has been trialled and outcome tested in ≥ 3 high quality RCTs indicating positive results

D. Information mean score = 4 *

* Exclude questions rated as "N/A" from the mean score calculation.

App subjective quality

SECTION E

20. **Would you recommend this app to people who might benefit from it?**

1 **Not at all** I would not recommend this app to anyone
2 There are very few people I would recommend this app to
3 **Maybe** There are several people whom I would recommend it to
4 There are many people I would recommend this app to
5 **Definitely** I would recommend this app to everyone

21. **How many times do you think you would use this app in the next 12 months if it was relevant to you?**

1 **None**
2 1-2
3 3-10
4 10-50
5 >50

22. **Would you pay for this app?**

1 No
3 Maybe
5 Yes

23. **What is your overall star rating of the app?**

1 ★ One of the worst apps I've used
2 ★★
3 ★★★ Average
4 ★★★★
5 ★★★★★ One of the best apps I've used

Scoring

App quality scores for

SECTION

A: Engagement Mean Score = 2.8
B: Functionality Mean Score = 4.25
C: Aesthetics Mean Score = 4
D: Information Mean Score = 4

App quality mean Score = 3.8

App subjective quality Score = 2

Mobile Application Rating Scale (MARS)

App Classification

The Classification section is used to collect descriptive and technical information about the app. Please review the app description in iTunes / Google Play to access this information.

App Name: T2 Mood Tracker

Rating this version: not enough ratings **Rating all versions:** (not enough ratings)

Developer: National Center for Telehealth & Technology

N ratings this version: - **N ratings all versions:** -

Version: 3.5.11 **Last update:** 2018

Cost - basic version: free **Cost - upgrade version:** -

Platform: ☐ iPhone ☐ iPad ☐ Android

Brief description: _____

Focus: what the app targets (select all that apply)

- ☐ Increase Happiness/Well-being
- ☐ Mindfulness/Meditation/Relaxation
- ☐ Reduce negative emotions
- ☐ Depression
- ☐ Anxiety/Stress
- ☐ Anger
- ☐ Behaviour Change
- ☐ Alcohol /Substance Use
- ☐ Goal Setting
- ☐ Entertainment
- ☐ Relationships
- ☐ Physical health
- ☐ Other Manage medication

Theoretical background/Strategies (all that apply)

- ☐ Assessment
- ☐ Feedback
- ☐ Information/Education
- ☐ Monitoring/Tracking
- ☐ Goal setting
- ☐ Advice /Tips /Strategies /Skills training
- ☐ CBT - Behavioural (positive events)
- ☐ CBT – Cognitive (thought challenging)
- ☐ ACT - Acceptance commitment therapy
- ☐ Mindfulness/Meditation
- ☐ Relaxation
- ☐ Gratitude
- ☐ Strengths based
- ☐ Other _____

Affiliations:

☐ Unknown ☐ Commercial ☐ Government ☐ NGO ☐ University

Age group (all that apply)

- ☐ Children (under 12)
- ☐ Adolescents (13-17)
- ☐ Young Adults (18-25)
- ☐ Adults
- ☐ General

Technical aspects of app (all that apply)

- ☐ Allows sharing (Facebook, Twitter, etc.)
- ☐ Has an app community
- ☐ Allows password protection
- ☐ Requires login
- ☐ Sends reminders
- ☐ Needs web access to function

App Quality Ratings

The Rating scale assesses app quality on four dimensions. All items are rated on a 5-point scale from "1.Inadequate" to "5.Excellent". Circle the number that most accurately represents the quality of the app component you are rating. Please use the descriptors provided for each response category.

SECTION A

Engagement – fun, interesting, customisable, interactive (e.g. sends alerts, messages, reminders, feedback, enables sharing), well-targeted to audience

1. **Entertainment: Is the app fun/entertaining to use? Does it use any strategies to increase engagement through entertainment (e.g. through gamification)?**
   1 Dull, not fun or entertaining at all
   2 Mostly boring
   3 OK, fun enough to entertain user for a brief time (< 5 minutes)
   4 Moderately fun and entertaining, would entertain user for some time (5-10 minutes total)
   5 Highly entertaining and fun, would stimulate repeat use

2. **Interest: Is the app interesting to use? Does it use any strategies to increase engagement by presenting its content in an interesting way?**
   1 Not interesting at all
   2 Mostly uninteresting
   3 OK, neither interesting nor uninteresting; would engage user for a brief time (< 5 minutes)
   4 Moderately interesting; would engage user for some time (5-10 minutes total)
   5 Very interesting, would engage user in repeat use

3. **Customisation: Does it provide/retain all necessary settings/preferences for apps features (e.g. sound, content, notifications, etc.)?**
   1 Does not allow any customisation or requires setting to be input every time
   2 Allows insufficient customisation limiting functions
   3 Allows basic customisation to function adequately
   4 Allows numerous options for customisation
   5 Allows complete tailoring to the individual's characteristics/preferences, retains all settings

4. **Interactivity: Does it allow user input, provide feedback, contain prompts (reminders, sharing options, notifications, etc.)? Note: these functions need to be customisable and not overwhelming in order to be perfect.**
   1 No interactive features and/or no response to user interaction
   2 Insufficient interactivity, or feedback, or user input options, limiting functions
   3 Basic interactive features to function adequately
   4 Offers a variety of interactive features/feedback/user input options
   5 Very high level of responsiveness through interactive features/feedback/user input options

5. **Target group: Is the app content (visual information, language, design) appropriate for your target audience?**
   1 Completely inappropriate/unclear/confusing
   2 Mostly inappropriate/unclear/confusing
   3 Acceptable but not targeted. May be inappropriate/unclear/confusing
   4 Well-targeted, with negligible issues
   5 Perfectly targeted, no issues found

A. Engagement mean score = 3.2

SECTION B

Functionality – app functioning, easy to learn, navigation, flow logic, and gestural design of app

6. **Performance: How accurately/fast do the app features (functions) and components (buttons/menus) work?**
   1 App is broken; no/insufficient/inaccurate response (e.g. crashes/bugs/broken features, etc.)
   2 Some functions work, but lagging or contains major technical problems
   3 App works overall. Some technical problems need fixing/Slow at times
   4 Mostly functional with minor/negligible problems
   5 Perfect/timely response; no technical bugs found/contains a 'loading time left' indicator

7. **Ease of use: How easy is it to learn how to use the app; how clear are the menu labels/icons and instructions?**
   1 No/limited instructions; menu labels/icons are confusing; complicated
   2 Useable after a lot of time/effort
   3 Useable after some time/effort
   4 Easy to learn how to use the app (or has clear instructions)
   5 Able to use app immediately; intuitive; simple

8. **Navigation: Is moving between screens logical/accurate/appropriate/ uninterrupted; are all necessary screen links present?**
   1 Different sections within the app seem logically disconnected and random/confusing/navigation is difficult
   2 Usable after a lot of time/effort
   3 Usable after some time/effort
   4 Easy to use or missing a negligible link
   5 Perfectly logical, easy, clear and intuitive screen flow throughout, or offers shortcuts

9. **Gestural design: Are interactions (taps/swipes/pinches/scrolls) consistent and intuitive across all components/screens?**
   1 Completely inconsistent/confusing
   2 Often inconsistent/confusing
   3 OK with some inconsistencies/confusing elements
   4 Mostly consistent/intuitive with negligible problems
   5 Perfectly consistent and intuitive

B. Functionality mean score = 4.5

SECTION C

Aesthetics – graphic design, overall visual appeal, colour scheme, and stylistic consistency

10. **Layout: Is arrangement and size of buttons/icons/menus/content on the screen appropriate or zoomable if needed?**
    1 Very bad design, cluttered, some options impossible to select/locate/see/read device display not optimised
    2 Bad design, random, unclear, some options difficult to select/locate/see/read
    3 Satisfactory, few problems with selecting/locating/seeing/reading items or with minor screen-size problems
    4 Mostly clear, able to select/locate/see/read items
    5 Professional, simple, clear, orderly, logically organised, device display optimised. Every design component has a purpose

11. **Graphics: How high is the quality/resolution of graphics used for buttons/icons/menus/content?**
    1 Graphics appear amateur, very poor visual design - disproportionate, completely stylistically inconsistent
    2 Low quality/low resolution graphics; low quality visual design – disproportionate, stylistically inconsistent
    3 Moderate quality graphics and visual design (generally consistent in style)
    4 High quality/resolution graphics and visual design – mostly proportionate, stylistically consistent
    5 Very high quality/resolution graphics and visual design - proportionate, stylistically consistent throughout

12. **Visual appeal: How good does the app look?**
    1 No visual appeal, unpleasant to look at, poorly designed, clashing/mismatched colours
    2 Little visual appeal – poorly designed, bad use of colour, visually boring
    3 Some visual appeal – average, neither pleasant nor unpleasant
    4 High level of visual appeal – seamless graphics – consistent and professionally designed
    5 As above + very attractive, memorable, stands out; use of colour enhances app features/menus

C. Aesthetics mean score = 3.7

SECTION D

Information – Contains high quality information (e.g. text, feedback, measures, references) from a credible source. Select N/A if the app component is irrelevant.

13. **Accuracy of app description (in app store): Does app contain what is described?**
    1 Misleading. App does not contain the described components/functions. Or has no description
    2 Inaccurate. App contains very few of the described components/functions
    3 OK. App contains some of the described components/functions
    4 Accurate. App contains most of the described components/functions
    5 Highly accurate description of the app components/functions

14. **Goals: Does app have specific, measurable and achievable goals (specified in app store description or within the app itself)?**
    N/A Description does not list goals, or app goals are irrelevant to research goal (e.g. using a game for educational purposes)
    1 App has no chance of achieving its stated goals
    2 Description lists some goals, but app has very little chance of achieving them
    3 OK. App has clear goals, which may be achievable.
    4 App has clearly specified goals, which are measurable and achievable
    5 App has specific and measurable goals, which are highly likely to be achieved

15. **Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?**
    N/A There is no information within the app
    1 Irrelevant/inappropriate/incoherent/incorrect
    2 Poor. Barely relevant/appropriate/coherent/may be incorrect
    3 Moderately relevant/appropriate/coherent/and appears correct
    4 Relevant/appropriate/coherent/correct
    5 Highly relevant, appropriate, coherent, and correct

16. **Quantity of information: Is the extent coverage within the scope of the app; and comprehensive but concise?**

N/A There is no information within the app
1 Minimal or overwhelming
2 Insufficient or possibly overwhelming
3 OK but not comprehensive or concise
4 Offers a broad range of information, has some gaps or unnecessary detail; or has no links to more information and resources
5 Comprehensive and concise; contains links to more information and resources

17. **Visual information: Is visual explanation of concepts – through charts/graphs/images/videos, etc. – clear, logical, correct?**

N/A There is no visual information within the app (e.g. it only contains audio, or text)
1 Completely unclear/confusing/wrong or necessary but missing
2 Mostly unclear/confusing/wrong
3 OK but often unclear/confusing/wrong
4 Mostly clear/logical/correct with negligible issues
5 Perfectly clear/logical/correct

18. **Credibility: Does the app come from a legitimate source (specified in app store description or within the app itself)?**

1 Source identified but legitimacy/trustworthiness of source is questionable (e.g. commercial business with vested interest)
2 Appears to come from a legitimate source, but it cannot be verified (e.g. has no webpage)
3 Developed by small NGO/institution (hospital/centre, etc.) /specialised commercial business, funding body
4 Developed by government, university or as above but larger in scale
5 Developed using nationally competitive government or research funding (e.g. Australian Research Council, NHMRC)

19. **Evidence base: Has the app been trialled/tested; must be verified by evidence (in published scientific literature)?**

N/A The app has not been trialled/tested
1 The evidence suggests the app does not work
2 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has partially positive outcomes in studies that are not randomised controlled trials (RCTs), or there is little or no contradictory evidence.
3 App has been trialled (e.g., acceptability, usability, satisfaction ratings) and has positive outcomes in studies that are not RCTs, and there is no contradictory evidence.
4 App has been trialled and outcome tested in 1-2 RCTs indicating positive results
5 App has been trialled and outcome tested in ≥ 3 high quality RCTs indicating positive results

D. Information mean score = 4.3 *

* Exclude questions rated as "N/A" from the mean score calculation.

App subjective quality

SECTION E

20. **Would you recommend this app to people who might benefit from it?**

1 **Not at all** I would not recommend this app to anyone
2 There are very few people I would recommend this app to
3 **Maybe** There are several people whom I would recommend it to
4 There are many people I would recommend this app to
5 **Definitely** I would recommend this app to everyone

21. **How many times do you think you would use this app in the next 12 months if it was relevant to you?**

1 **None**
2 1-2
3 3-10
4 10-50
5 >50

22. **Would you pay for this app?**

1 No
3 Maybe
5 Yes

23. **What is your overall star rating of the app?**

1 ★ One of the worst apps I've used
2 ★★
3 ★★★ Average
4 ★★★★
5 ★★★★★ One of the best apps I've used

Scoring

App quality scores for

SECTION

A: Engagement Mean Score = 3.2

B: Functionality Mean Score = 4.5

C: Aesthetics Mean Score = 3.7

D: Information Mean Score = 4.3

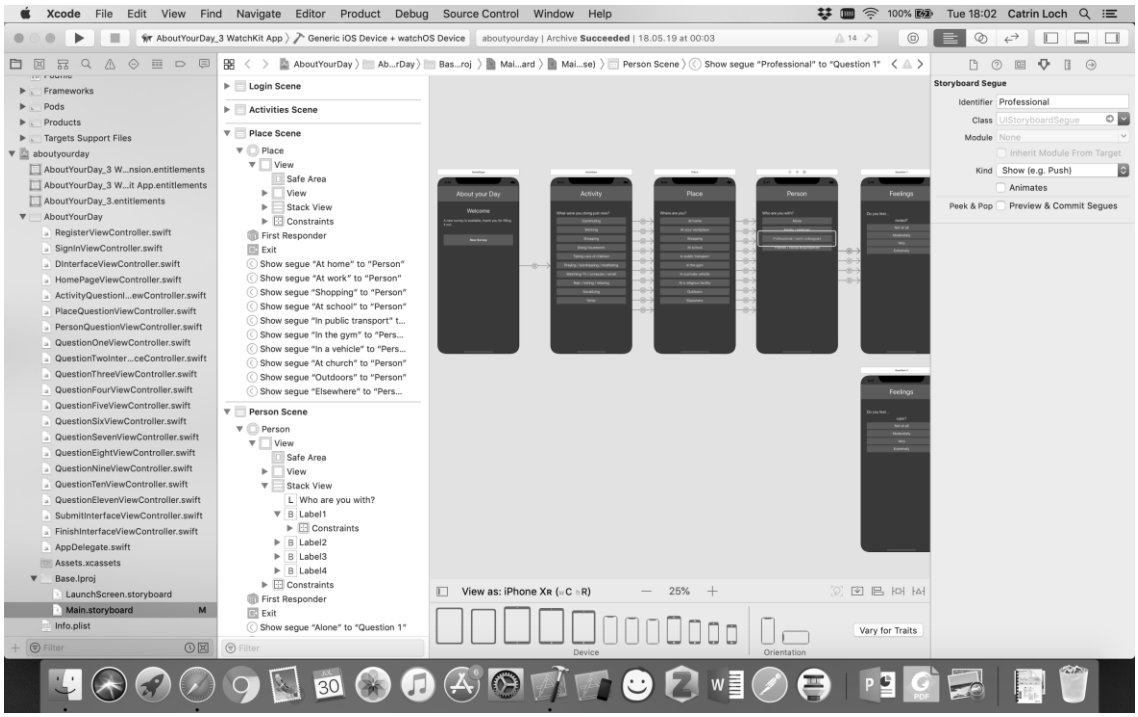
App quality mean Score = 3.9

App subjective quality Score = 1.75

Requirement: Present Assessment Content to Subjects

XCode provides the developer with so called storyboards, where the developer can manage his different views and interfaces for the smartphone (see Picture below) and smartwatch application. By dragging and dropping the developer can add the UI pieces from the library he wants to design his view with, for example buttons, labels or titles. Furthermore, the developer can link multiple views with each other by adding an action to a button and linking that action to the next view. For example, the button “Professional / work colleague” in the segue “Person” as been assigned to push the view “Feelings” if it has been pressed by the user (see Picture below). In order to design for multiple Apple devices, XCode also provides the possibility to switch between different devices in the storyboard. Moreover, this did helps to consider the different sizes of the screens and adjust the board constraints if necessary.

More specific, for the feeling questions the labels and buttons were only used as placeholder, which were overwritten with the content from the array. If the random assignment would not work by accident, then the users would still have the possibility to fill in the survey.

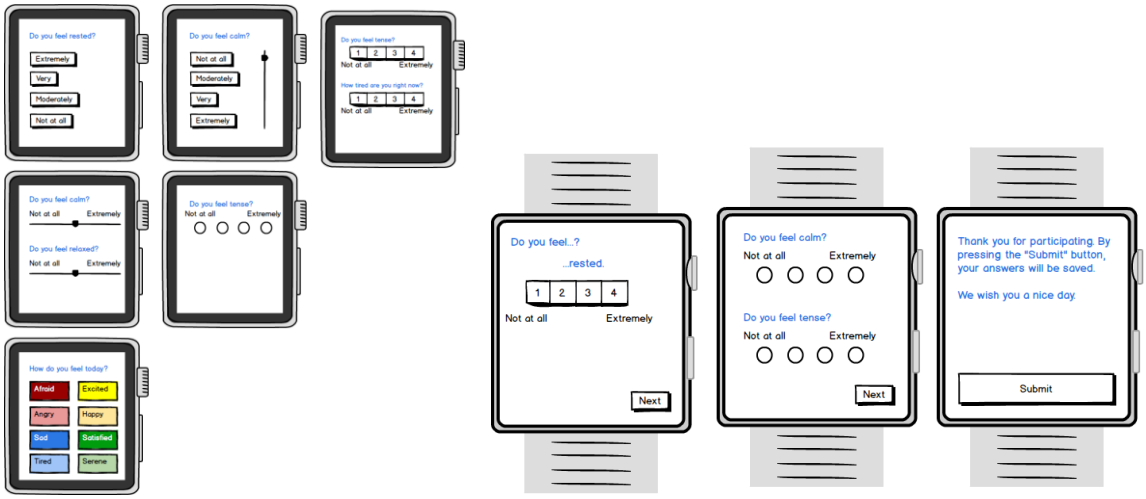


About your Day: App Description

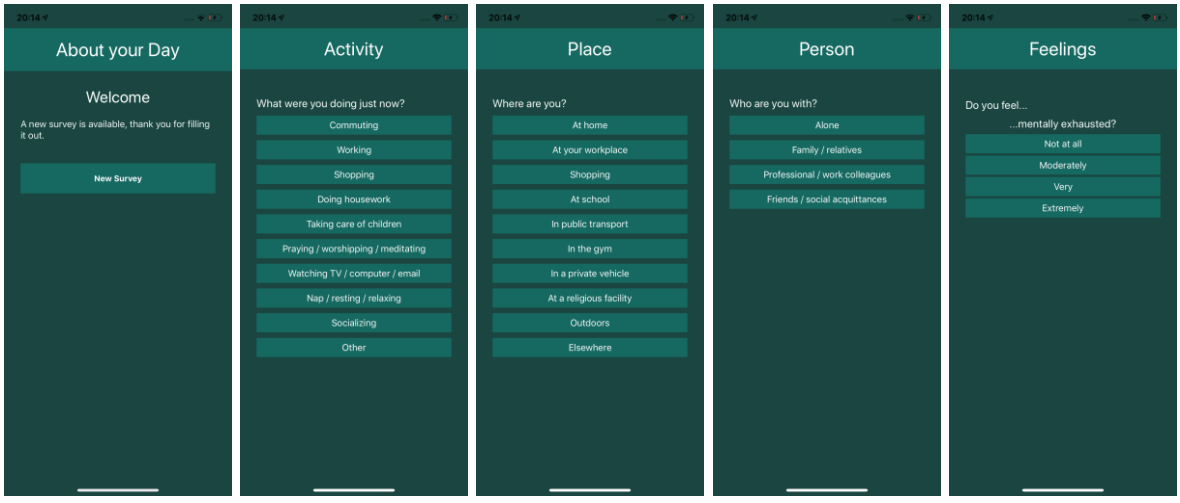
An iOS and Apple Watch survey tool for the «About your Day» study. An exploratory study about self-reporting methods and people’s responses to these different methods. It is hosted by researchers at the University of Zurich, Switzerland in agreement with ethical regulations of this institution.

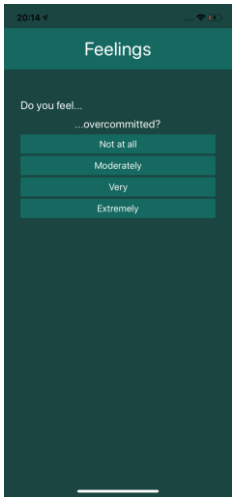
Prototype Iterations Smartphone

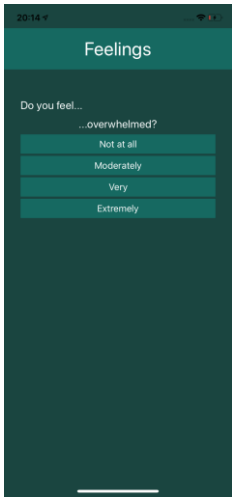
Prototype Iterations Smartwatch

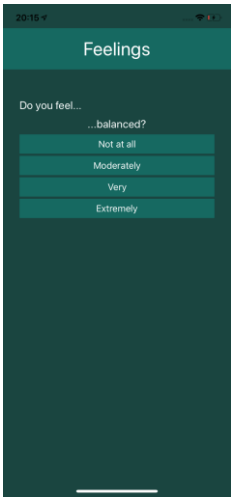


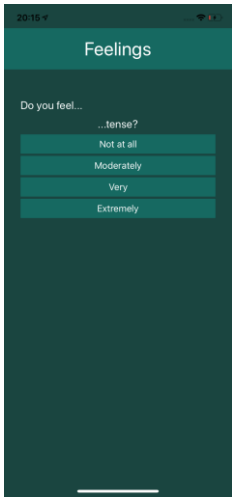
Prototype Smartphone

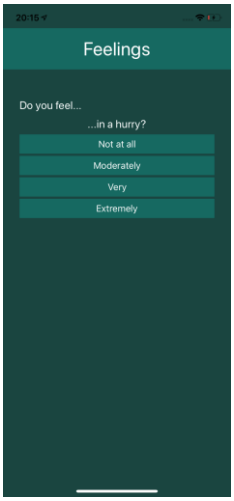


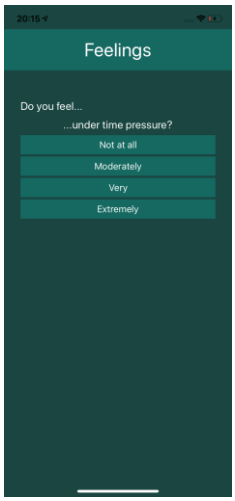

Feelings
Do you feel...
...overcommitted?
Not at all
Moderately
Very
Extremely

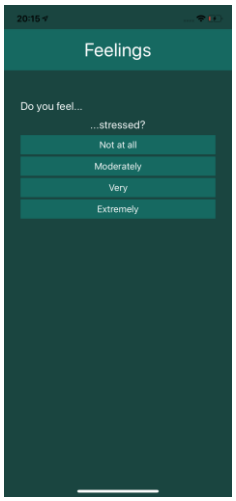

Feelings
Do you feel...
...overwhelmed?
Not at all
Moderately
Very
Extremely

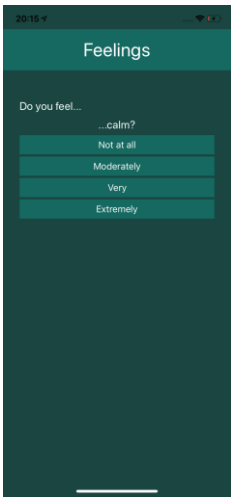

Feelings
Do you feel...
...balanced?
Not at all
Moderately
Very
Extremely

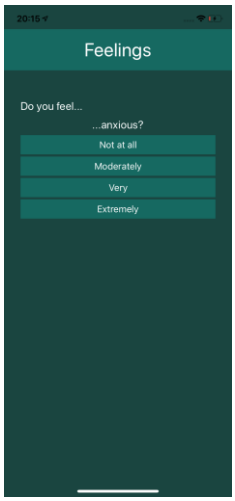

Feelings
Do you feel...
...tense?
Not at all
Moderately
Very
Extremely

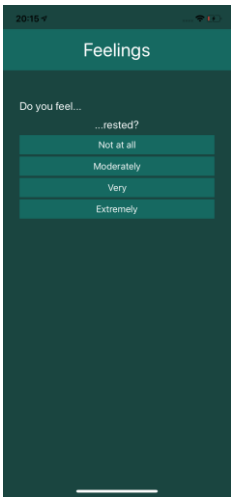

Feelings
Do you feel...
...in a hurry?
Not at all
Moderately
Very
Extremely

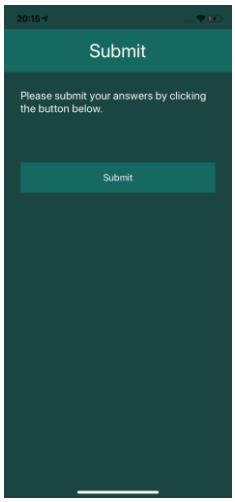

Feelings
Do you feel...
...under time pressure?
Not at all
Moderately
Very
Extremely

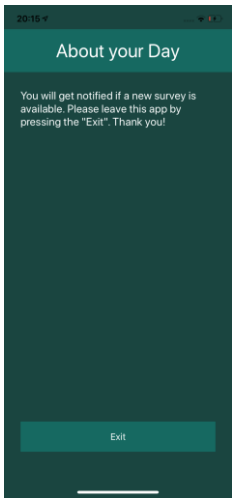

Feelings
Do you feel...
...stressed?
Not at all
Moderately
Very
Extremely


Feelings
Do you feel...
...calm?
Not at all
Moderately
Very
Extremely

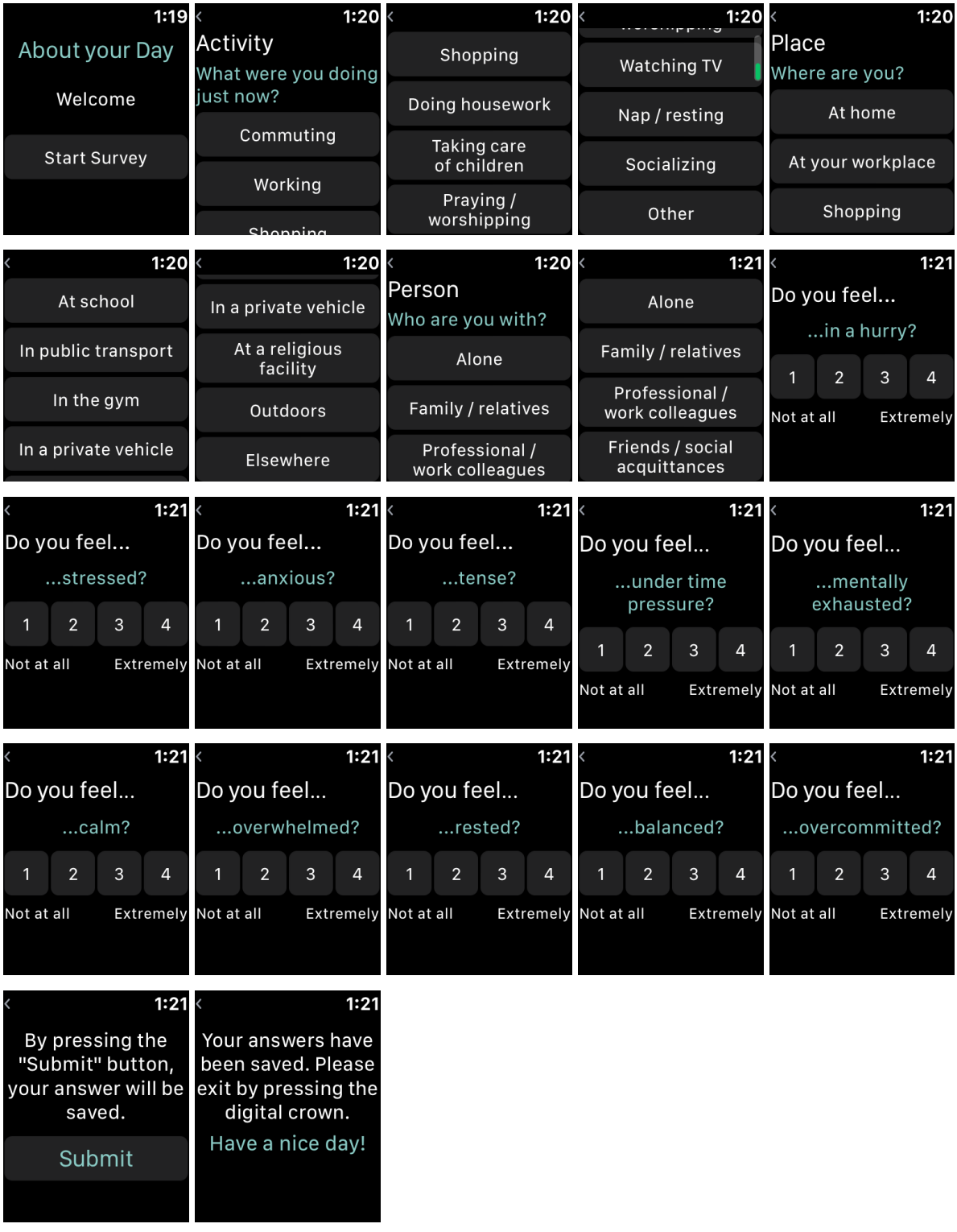

Feelings
Do you feel...
...anxious?
Not at all
Moderately
Very
Extremely


Feelings
Do you feel...
...rested?
Not at all
Moderately
Very
Extremely

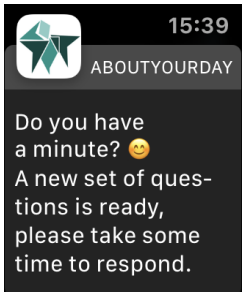

Submit
Please submit your answers by clicking the button below.
Submit

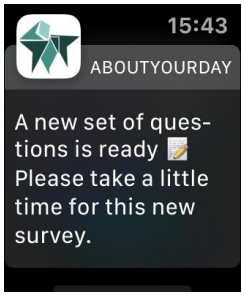

About your Day
You will get notified if a new survey is available. Please leave this app by pressing the "Exit". Thank you!
Exit

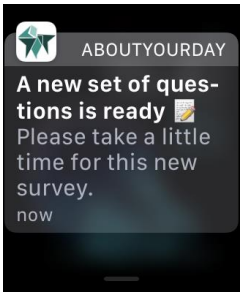
Prototype Smartwatch



Notification on Apple Watch







Push Notification Schedule

Date	Time	Notification Title	Notification Text	Device
20.05.2019	8.28	Good Morning	If you have a minute, please fill out the new survey.	iPhone
	10.24	A new set of questions is ready	Please take a little time for this new survey.	
	12.11	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.24	A new set of questions is ready	Take a little time for this new survey.	
	18.54	Good Evening	If you have a minute, please fill out the new survey.	
20.05.2019	10.19	Good Morning	If you have a minute, please fill out the new survey.	Watch
	11.58	A new set of questions is ready	Take a little time for this new survey.	
	13.50	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	16.45	A new set of questions is ready	Take a little time for this new survey.	
	18.15	Good Evening	If you have a minute, please fill out the new survey.	
21.05.2019	9.42	Good Morning	If you have a minute, please fill out the new survey.	iPhone
	10.29	A new set of questions is ready	Take a little time for this new survey.	
	12.59	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	14.45	A new set of questions is ready	Take a little time for this new survey.	
	17.45	Good Evening	If you have a minute, please fill out the new survey.	
21.05.2019	09.54	Good Morning	If you have a minute, please fill out the new survey.	Watch
	10.16	A new set of questions is ready	Take a little time for this new survey.	
	14.51	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.33	A new set of questions is ready	Take a little time for this new survey.	
	18.56	Good Evening	If you have a minute, please fill out the new survey.	
22.05.2019	10.19	Good Morning	If you have a minute, please fill out the new survey.	iPhone
	11.23	A new set of questions is ready	Take a little time for this new survey.	
	12.48	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.40	A new set of questions is ready	Take a little time for this new survey.	
	16.45	Do you have a minute?	If you have a minute, please fill out the new survey.	
22.05.2019	09.15	Good Morning	If you have a minute, please fill out the new survey.	Watch
	10.29	A new set of questions is ready	Take a little time for this new survey.	
	12.37	Do you have a minute?	A new set of questions is ready, please take some time to respond.	

	14.31	A new set of questions is ready	Take a little time for this new survey.	
	16.43	Do you have a minute?	If you have a minute, please fill out the new survey.	
23.05.2019	8.28	Good Morning	If you have a minute, please fill out the new survey.	Phone
	10.24	A new set of questions is ready	Take a little time for this new survey.	
	12.11	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.24	A new set of questions is ready	Take a little time for this new survey.	
	18.54	Good Evening	If you have a minute, please fill out the new survey.	
23.05.2019	10.19	Good Morning	If you have a minute, please fill out the new survey.	Watch
	11.58	A new set of questions is ready	Take a little time for this new survey.	
	13.50	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	16.45	A new set of questions is ready	Take a little time for this new survey.	
	18.15	Good Evening	If you have a minute, please fill out the new survey.	
24.05.2019	8.42	Good Morning	If you have a minute, please fill out the new survey.	Phone
	10.29	A new set of questions is ready	Take a little time for this new survey.	
	12.59	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	14.45	A new set of questions is ready	Please take a little time for this new survey.	
	17.45	Good Evening	If you have a minute, please fill out the new survey.	
24.05.2019	08.54	Good Morning	If you have a minute, please fill out the new survey.	Watch
	10.16	A new set of questions is ready	Take a little time for this new survey.	
	14.51	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.33	A new set of questions is ready	Take a little time for this new survey.	
	18.56	Good Evening	If you have a minute, please fill out the new survey.	
25.05.2019	09.19	Good Morning	If you have a minute, please fill out the new survey.	Phone
	11.23	A new set of questions is ready	Take a little time for this new survey.	
	12.48	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.40	A new set of questions is ready	Take a little time for this new survey.	
	17.45	Good Evening	If you have a minute, please fill out the new survey.	
25.05.2019	09.15	Good Morning	If you have a minute, please fill out the new survey.	Watch
	10.29	A new set of questions is ready	Please take a little time for this new survey.	
	12.37	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	14.31	A new set of questions is ready	Please take a little time for this new survey.	
	17.43	Good Evening	If you have a minute, please fill out the new survey.	
26.05.2019	9.00	Good Morning	If you have a minute, please fill out the new survey.	Phone
	10.24	A new set of questions is ready	Please take a little time for this new survey.	
	12.11	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.24	A new set of questions is ready	Please take a little time for this new survey.	
	18.54	Good Evening	If you have a minute, please fill out the new survey.	
26.05.2019	10.19	Good Morning	If you have a minute, please fill out the new survey.	Watch
	11.58	A new set of questions is ready	Please take a little time for this new survey.	

	13.50	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	16.45	A new set of questions is ready	Please take a little time for this new survey.	
	18.15	Good Evening	If you have a minute, please fill out the new survey.	
27.05.2019	8.42	Good Morning	If you have a minute, please fill out the new survey.	Phone
	10.29	A new set of questions is ready	Please take a little time for this new survey.	
	12.59	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	14.45	A new set of questions is ready	Please take a little time for this new survey.	
	17.45	Good Evening	If you have a minute, please fill out the new survey.	
27.05.2019	08.54	Good Morning	If you have a minute, please fill out the new survey.	Watch
	10.16	A new set of questions is ready	Please take a little time for this new survey.	
	13.51	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.33	A new set of questions is ready	Please take a little time for this new survey.	
	18.56	Good Evening	If you have a minute, please fill out the new survey.	
28.05.2019	10.19	Good Morning	If you have a minute, please fill out the new survey.	Phone
	11.23	A new set of questions is ready	Please take a little time for this new survey.	
	12.48	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	15.40	A new set of questions is ready	Please take a little time for this new survey.	
	16.45	Do you have a minute?	Please fill out the new survey.	
28.05.2019	09.15	Good Morning	If you have a minute, please fill out the new survey.	Watch
	11.29	A new set of questions is ready	Please take a little time for this new survey.	
	12.37	Do you have a minute?	A new set of questions is ready, please take some time to respond.	
	14.31	A new set of questions is ready	Please take a little time for this new survey.	
	16.43	Do you have a minute?	Please fill out the new survey.	

Push Notification Template Text

Title	**Text**
Good Morning	If you have a minute, please fill out the new survey.
A new set of questions is ready 📝	Please take a little time for this new survey.
Do you have a minute? 😊	A new set of questions is ready, please take some time to respond.
A new set of questions is ready 📝	Take a little time for this new survey.
Good Evening	If you have a minute, please fill out the new survey.

About your Day: Links & Credentials

LimeSurvey: https://aboutyourday.ifi.uzh.ch/limesurvey/admin

Username: admin
Password: a9R3Vpi2
Language: Default
Access to:

- Demographic Questionnaire
- Computer Survey for Day 1 to 3
- Feedback Questionnaire Computer, Smartphone and Smartwatch
- Overall Feedback Questionnaire
- Collected Data with LimeSurvey

Sever: https://aboutyourday.ifi.uzh.ch/phpmyadmin/

Username: eva
Password: UC0thoPoWith
Access to:

- Smartphone data table: phonedata
- Smartwatch data table: watchdata
- Heart Rate data table: heartratedata
- Collected Data with LimeSurvey

OneSignal for Push Notifications: https://onesignal.com/

Username: lghtrs@gmail.com
Password: please contact Catrin Loch
Access to:

- Managing Push Notifications
- Statistics regarding who opened notifications
- Participants devices

About your Day Application Source Code:
https://gitlab.ifi.uzh.ch/ZPAC/exploring-methods-for-self-reporting-stress/tree/develop

Appendix B: About your Day Study Content

Schedule for Participants

1. Register for Participation by sending an email with name, occupation and versions of devices
2. Get information on study
3. Accept Informed consent
4. Demographic Questionnaire
5. Download Survey Application
6. Receive study schedule
7. Self-reporting with first condition
8. Feedback Questionnaire to first condition

9. Self-reporting with second condition
10. Feedback Questionnaire to second condition
11. Self-reporting with third method
12. Feedback Questionnaire to third condition
13. Overall Feedback Questionnaire
14. Debriefing including Fact Sheet and Compensation

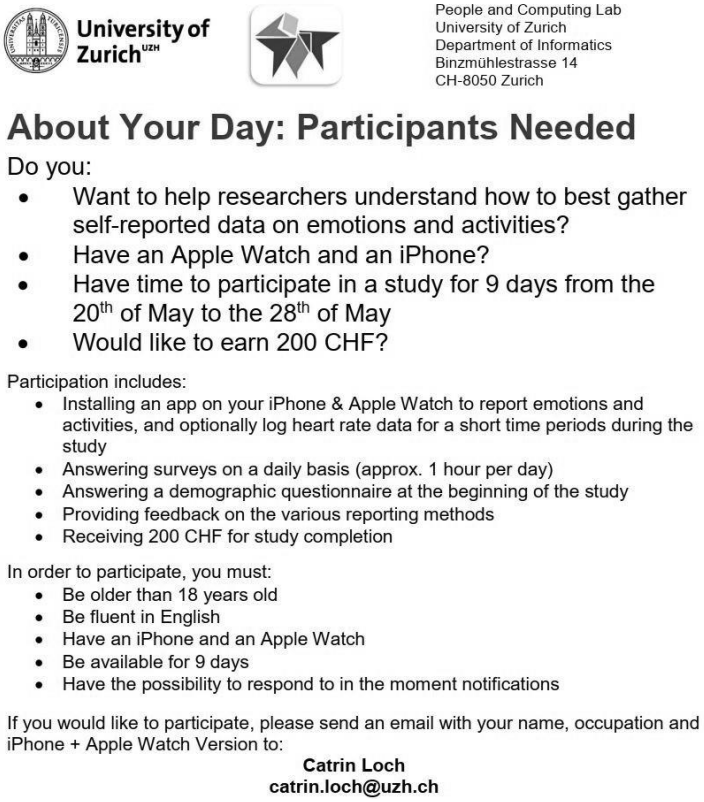
Demographic Questionnaire

1. What is your gender? Female Male Other Prefer not to say
2. In what year were you born?
3. Are you…? Please select all that apply.
   - □ Employed or self-employed
   - □ Studying or in school
   - □ Retired
   - □ Unemployed and seeking work
   - □ Long-term sick or disabled
   - □ Looking after family or home
4. Who else lives in your household? Please select all that apply.
   - □ I live alone
   - □ Spouse / partner
   - □ My children
   - □ Flatmate(s)
   - □ Parent(s)
   - □ Relatives other than parents
5. How satisfied are you across all aspects of your life in the past year?
   - □ Not at all satisfied
   - □ Not very satisfied
   - □ Satisfied
   - □ Very satisfied

   How satisfied are you with your personal life over the past year?
   - □ Not at all satisfied
   - □ Not very satisfied
   - □ Satisfied
   - □ Very satisfied

   How satisfied are you with your professional life or job over the past year?
   - □ Not at all satisfied
   - □ Not very satisfied
   - □ Satisfied
   - □ Very satisfied

Recruiting Flyer



University of Zurich[UZH]

People and Computing Lab
University of Zurich
Department of Informatics
Binzmühlestrasse 14
CH-8050 Zurich

About Your Day: Participants Needed

Do you:

- Want to help researchers understand how to best gather self-reported data on emotions and activities?
- Have an Apple Watch and an iPhone?
- Have time to participate in a study for 9 days from the 20th of May to the 28th of May
- Would like to earn 200 CHF?

Participation includes:

- Installing an app on your iPhone & Apple Watch to report emotions and activities, and optionally log heart rate data for a short time periods during the study
- Answering surveys on a daily basis (approx. 1 hour per day)
- Answering a demographic questionnaire at the beginning of the study
- Providing feedback on the various reporting methods
- Receiving 200 CHF for study completion

In order to participate, you must:

- Be older than 18 years old
- Be fluent in English
- Have an iPhone and an Apple Watch
- Be available for 9 days
- Have the possibility to respond to in the moment notifications

If you would like to participate, please send an email with your name, occupation and iPhone + Apple Watch Version to:

Catrin Loch
catrin.loch@uzh.ch

Consent Form

University of Zurich[UZH]

People and Computing Lab
University of Zurich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zurich

Contact person:

Catrin Loch
Phone +41 76 358 75 78
catrin.loch@uzh.ch
Prof. Dr. Thomas Fritz
fritz@ifi.uzh.ch
University of Zurich

Jan Gugler
jan.gugler@uzh.ch
University of Zurich
Prof. Dr. Elaine Huang
huang@ifi.uzh.ch
University of Zurich

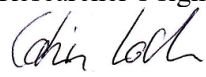
Informed Consent Form

A pilot study to explore methods of self-reporting activities and emotions

Dear Sir or Madam,
In this study, we investigate different methods of self-reporting activities and emotions and assess people's response to these methods. By participating, you contribute to research intended to increase our understanding of different self-reporting methods. It is hosted by researchers at the University of Zurich, Switzerland in agreement with ethical regulations of this institution.

What will I be asked to do?
If you choose to participate in this study, you will be asked to:
1) answer a demographic questionnaire
2) fill out surveys about your feelings and mood as requested
3) fill out feedback questionnaires
The total duration of this study is 9 days, during which time you can expect to spend on average 60 minutes per day participating actively.
What information will be collected?
You will be asked to fill out the demographic questionnaire to provide some insights about your personal background (e.g. age, life satisfaction). Additionally, you will be asked to answer questions regarding your activities and feelings during the day. If you are willing, we will also collect your heart rate measures via Apple Watch for short periods of time.
What happens to the data?
All of your original data will be saved on password-protected devices, made anonymous and will be treated as confidential. Your data will also be stored for six months after finishing the data analysis.
Are there risks to participating?
There are no risks of participating in this study. Participation in the study is voluntary and confidential. Your data will be anonymized. If it is ever shared with anyone outside of the research team, including in any written publications or oral presentations based on this research, you will be identified only by a participant number (e.g. P03) or a pseudonym.
You are free to withdraw from participation at any point during the study without needing to provide any reasons. However, unless you request otherwise, any information you contribute up to the point at which you choose to withdraw will be retained and may be used in the study.
Are there any benefits to participating?
You will help advance the scientific understanding about self-reporting methods and your participation will be compensated with 200 CHF. After the study, if you are interested in our research results or participating in any future studies on this topic, we will be happy to keep you informed.
Consent
By signing this form, you confirm the following statements:
• A researcher explained the study and the listed conditions to me.
• I had the opportunity to ask questions.
• I understood the answers and accept them.
• I am at least 18 years old.
• I had enough time to make the decision to participate.
• I agree to the participation.
• I permit the use of my anonymized data by other researchers affiliated with the University of Zurich People and Computing Lab for further study and analysis.
• I permit the use of my anonymized data for internal or external publications and presentations to this research.

In no way does signing this form waive your legal rights or release the investigators or involved institutions from their legal or professional responsibilities. You are free to withdraw from this research project at any time. Please feel free to ask for clarification or new information at any time during your participation.

Participant's name	Researcher's name
_____	_____
Location and date	Location and date
_____	_____
Participant's signature	Researcher's signature
_____	

Questions or Concerns?
A copy of this consent form has been given to you to keep. The researcher has kept a copy of the consent form. If you have further questions regarding our research, and/or your participation in this study, please contact:

Catrin Loch
catrin.loch@uzh.ch
University of Zurich

Prof. Dr. Elaine Huang
huang@ifi.uzh.ch
University of Zurich

Prof. Dr. Thomas Fritz
fritz@ifi.uzh.ch
University of Zurich

Counterbalancing Table

Counterbalancing Table	20.05.19	21.05.19	22.05.19	23.05.19	24.05.19	25.05.19	26.05.19	27.05.19	28.05.19
	Mo	**Di**	**Mi**	**Do**	**Fr**	**Sa**	**So**	**Mo**	**Di**
P01	Daily	Daily	Daily	Phone	Phone	Phone	Watch	Watch	Watch
P02	Daily	Daily	Daily	Watch	Watch	Watch	Phone	Phone	Phone
P03	Phone	Phone	Phone	Watch	Watch	Watch	Daily	Daily	Daily
P04	Phone	Phone	Phone	Daily	Daily	Daily	Watch	Watch	Watch
P05	Watch	Watch	Watch	Phone	Phone	Phone	Daily	Daily	Daily
P06	Watch	Watch	Watch	Daily	Daily	Daily	Phone	Phone	Phone
P07	Daily	Daily	Daily	Phone	Phone	Phone	Watch	Watch	Watch
P08	Watch	Watch	Watch	Phone	Phone	Phone	Daily	Daily	Daily
P09	Phone	Phone	Phone	Daily	Daily	Daily	Watch	Watch	Watch
P10	Daily	Daily	Daily	Watch	Watch	Watch	Phone	Phone	Phone
P11	Phone	Phone	Phone	Watch	Watch	Watch	Daily	Daily	Daily
P12	Watch	Watch	Watch	Daily	Daily	Daily	Phone	Phone	Phone

Participants Schedule

People and Computing Lab
University of Zurich
Department of Informatics
Binzmühlestrasse 14
CH-8050 Zurich

Contact person:
Catrin Loch
Phone +41 76 358 75 78
catrin.loch@uzh.ch
Prof. Dr. Thomas Fritz
fritz@ifi.uzh.ch
University of Zurich

Prof. Dr. Elaine Huang
huang@ifi.uzh.ch
University of Zurich

About your Day: Participant Schedule

Dear <Participant Name>

Thank you for being part of this study and taking your time to participate. The next few days contain the following tasks for you:

20.05.-22.05.19: Computer Questionnaire

In the morning of the 20th of May we are going to send you a reminder via email. In your first three days you are going to receive an invitation to fill out a computer questionnaire once per day. Please use your preferred computer browser on your computer to fill in your answers. After these three days we are going to send you an invitation for a feedback questionnaire.

23.05.-25.05.19: Watch Questionnaires

In the morning of the 23rd of May we are going to send you a reminder that the phone surveys are over, and you have to switch to your Apple Watch. The next three days you are going to be notified on your phone or Apple Watch if there is an available survey, please fill it in on your Apple Watch and do not forget to exit the application by pressing the digital crown. You can directly click into the Apple Watch Notification to open the questionnaire. Please make sure that you have internet access as often as you can and reply to the notifications as soon as you can. If any problems occur, please fully close the application and wait for the next notification. After these three days we are going to send you an invitation for a feedback questionnaire, please fill it out as soon as you can.

26.05.-28.05.19: Phone Questionnaires

In the morning of the 26th of May we are going to send you a reminder that the phone surveys are over, and you have to switch to your iPhone. The next three days you are going to be notified on your phone or Apple Watch if there is an available survey, please fill it in on your iPhone and do not forget to exit the application. Just open the notification on your iPhone and you will be directed to the survey. If any problems occur, please fully close the application and wait for the next notification. After these three days we are going to send you an invitation for a feedback questionnaire.

28.05.19: Final Feedback Questionnaire

On the 28th of May you are going to receive an email with the invitation for a final feedback questionnaire. After filling out this overall feedback questionnaire, the study is over.

After the study

In the next few days after the study, the researcher team are going to get in contact with you to organize the compensation. If you are interested in finding out more on your emotions and feelings during these days, please let us know and we will provide you with a small fact sheet.

In general, if you miss a notification or close it per accident, no worries a new one will appear. If you have any questions, please do not hesitate to contact me.

Best wishes,

Catrin & the Research Team

Examiner Schedule

1. Mailing with Flyer and hang up Flyer; specifically ask people to participate
2. Email with Consent Form as pdf and Link to the demographic questionnaire
3. Ask participants for UUID number of the device → provide them with short guide how to
4. Provide participants with schedule + Invitation to Application
5. Create participant tables in LimeSurvey for each questionnaire
6. Adjust email templates Limesurvey in every questionnaire
7. Create segments in one signal for participant groups

Day 1

1. Reminder Email that study starts
2. Check if notifications have been sent out and people filled out questionnaires
3. 17.30 send email with computer survey, availability until midnight
4. Remove access if computer has not been filled out until midnight

---

Day 2

1. Check if notifications have been sent out and people filled out questionnaires
2. 17.30 send email with computer survey, availability until midnight
3. Remove access if computer has not been filled out until midnight

---

Day 3

1. Check if notifications have been sent out and people filled out questionnaires
2. 17.30 send email with computer survey, availability until midnight
3. Remove access if computer has not been filled out until midnight
4. Day 3: send Feedback Survey Link after the last trigger, around 8pm

---

Day 4

1. Check if notifications have been sent out and people filled out questionnaires
2. 12.00 Uhr reminder feedback questionnaire
3. 17.30 send email with computer survey, availability until midnight
4. Remove access if computer has not been filled out until midnight

---

Day 5

1. Check if notifications have been sent out and people filled out questionnaires
2. 17.30 send email with computer survey, availability until midnight
3. Remove access if computer has not been filled out until midnight

---

Day 6

1. Check if notifications have been sent out and people filled out questionnaires
2. 17.30 send email with computer survey, availability until midnight
3. Remove access if computer has not been filled out until midnight
4. Day 3: send Feedback Survey Link after the last trigger, around 8pm

---

Day 7

1. Check if notifications have been sent out and people filled out questionnaires
2. 12.00 Uhr Reminder feedback questionnaire
3. 17.30 send email with computer survey, availability until midnight
4. Remove access if computer has not been filled out until midnight

---

Day 8

1. Check if notifications have been sent out and people filled out questionnaires
2. 17.30 send email with computer survey, availability until midnight
3. Remove access if computer has not been filled out until midnight

---

Day 9

1. Check if notifications have been sent out and people filled out questionnaires
2. 17.30 send email with computer survey, availability until midnight
3. Remove access if computer has not been filled out until midnight
4. Day 3: send Feedback Survey Link after the last trigger, around 8pm
5. Send overall feedback survey → 3 days availability

Day 10-12

1. Tag 11: 12.00 Uhr Reminder Feedback questionnaire
2. Provide fact sheets to participants
3. Send thank you email +Get in contact with participants for compensation

Make Application Downloadable

The application has been made available for the participants on TestFlight Beta Testing which is an Apple product on which beta versions can be made available. Testers can be added as internal testers or external testers. As the application has go to through an extensive review by Apple when using external testers, we decided to add participants as internal testers. Internal testers are iTunes Connect users with an Admin, App Manager, Legal, Developer or Marketer role with access to the app.

Steps to make app downloadable:

In Xcode:

1. Choose Generic iOS Device in the scheme chooser
2. Product > Archive, which opens the Organizer window with your app in the archives tab
3. Click Upload to App Store, go through the process and click upload
4. The app is now being uploaded to iTunes Connect and is now available on iTunes Connect

In iTunes Connect:

5. Add participants as internal tester, by heading to the Users and Roles section in iTunes Connect
6. Click the + button to add a new user, fill in the user info with an email address and click next. The email address needs to be associated to an AppleID otherwise the user needs to create a new AppleID.
7. For this specific case, participants were added with a Customer Support role, which had the least privileges. iTunes Connect sends an invitation to the new user and the user needs to verify his or her account. (See steps how to download)
8. If the user has accepted the invitation, go to the My Apps section and click on your app
9. Select the TestFlight tab and click “Add iTunes Connect Users” in the left side menu. Select the internal testers you’d like to add as internal testers for this build and click Add.
10. All selected testers receive and email with a link to download and install this app via the TestFlight app.

Application Download Description for Participant

People and Computing Lab
University of Zurich
Department of Informatics
Binzmühlestrasse 14
CH-8050 Zurich

Contact person:
Catrin Loch
Phone +41 76 358 75 78
catrin.loch@uzh.ch

Prof. Dr. Thomas Fritz
fritz@ifi.uzh.ch
University of Zurich

Prof. Dr. Elaine Huang
huang@ifi.uzh.ch
University of Zurich

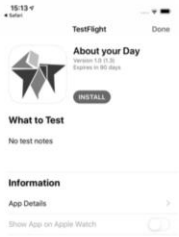
About your Day: Download Study Application

This procedure takes approx. 30 minutes, please download it as soon as you can.
Remark: Downloading the About your Day application entails to create an App Store Connect Login which is associated to your Apple ID. If you have successfully gone through the steps 1)-4) your name will be visible for other test users associated to the project. After informing Catrin in step 17) your login can be removed and your name will not be visible anymore. If you have any concerns about this, please inform Catrin in case you do not want to participate anymore.
Thank you for your understanding.

1) Go to your inbox and find the email entitled "App Store Connect" with the invitation of Catrin "You've been invited to App Store Connect"
2) Open your invitation email on your iOS device and click "Accept invitation"
3) You arrive at the App Store Connect Webpage
4) Enter your AppleID and password and verify your account
5) Inform Catrin that you have gone through all these steps and she will then add you as a test user
6) Install the TestFlight app on your iPhone

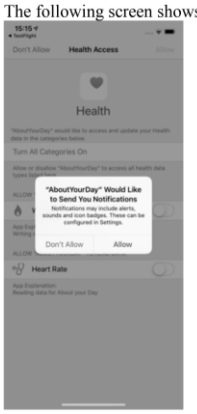
7) You receive a new email entitled "Catrin Loch has invited you to test About your Day"
8) Open the email on your iPhone and click on "View in TestFlight"
9) Follow the steps in the popup
10) If everything worked, you should end up on the screen in step 11)

11) If the application is not downloaded automatically, you should end up on the following screen. Click on install and wait until the application is installed.

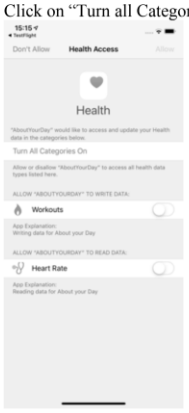


12) When the installation is finished you see the following screen. Please Check that "Installing on Apple Watch.." is accepted and that your screen looks like the following screenshot. Then click on "Done".

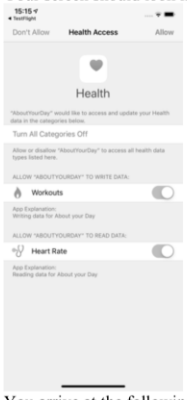
13) The following screen shows up. Click on “Allow” Notifications and the popup disappears.



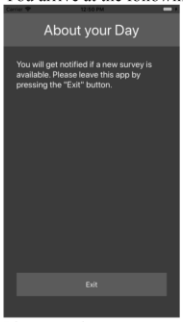
14) Click on “Turn all Categories On” in the Health App



15) Your screen should look like the following. Click on “Allow”.



16) You arrive at the following screen. Please read through the message and go to the final step.



17) Please send a message to Catrin (catrin.loch@uzh.ch, 076 358 75 78) when you finished installing the application. Thank you!

Computer Questionnaire

We would like to learn what you did and how you felt today. Not all days are the same – some are better, some are worse and others are pretty typical. Here we are only asking you about today. Because many people find it difficult to remember what exactly they did and experienced, we will do this in two steps:

1. On this page, we will ask you when you woke up and when you went to sleep yesterday.
2. We'd like you to reconstruct what your day was like, as if you were writing in your diary. Where were you? What did you do and experience? How did you feel? Answering the questions on the next page will help you to reconstruct your day.

To begin, please select the day of the week that today was:

Monday Tuesday Wednesday Thursday Friday Saturday Sunday

About what time did you wake up?

Before 6am
Between 6am and 7am
Between 7am and 8am
Between 8am and 9am
Between 9am and 10am
After 10am

And when did you go to sleep yesterday?

Before 7pm
Between 7pm and 8pm
Between 8pm and 9pm
Between 9pm and 10pm
Between 10pm and 11pm
After 11pm

This Morning (8am to 11am)

What were you doing this morning? (Check all that apply)

Commuting
Working
Shopping
Doing housework
Taking care of your children
Praying / worshipping / meditating
Watching TV / computer / email / internet
Nap / resting / relaxing
Socializing
Other

Where were you?

At home
At work

Shopping
At school
In public transport
In the gym
In a private vehicle
At a religious facility
Outdoors
Elsewhere

Where you with anyone?
Yes No
If yes, with whom did you spend the most time?
Family / relatives
Professional / work colleagues
Friends / social acquittances

How did you feel in the morning?
Please rate each feeling on the scale given. A rating of 1 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please check the number between 1 and 6 that best describes how you felt.

In this situation, did you feel…?	Not at all 1	2	3	4	5	Very much 6
Rested						
Calm						
Tense						
In a hurry						
Mentally exhausted						
Overwhelmed						
Overcommitted						
Anxious						
Balanced						
Stressed						
Under time pressure						

The same questions as above are answered for the midday (11am to 2pm) episode and afternoon (2pm to 5pm) episode as well.

Feedback Questionnaire Computer

- On a scale from 1 to 5, how disruptive was it for you to fill out the questionnaire at the end of the day?
  1= not at all, 5 = totally
- On a scale from 1 to 5, how much do you feel your environment (e.g. the people around you) influenced the way you answered the questions?
  1= not at all, 5 = totally
- On a scale from 1 to 5, do you think the necessary to complete a single survey was accurate ?
  1= not at all, 5 = very

- On a scale from 1 to 5, how burdensome was it to fill out the survey?
  1 = not at all, 5 = very
- On a scale from 1 to 5, do you think the questions were difficult to understand?
  1 = not at all, 5 = very
- On a scale from 1 to 5, how easy was it to complete the survey?
  1 = not at all, 5 = very
- On a scale from 1 to 5, did the form of the survey, i.e. that you answered the questions on your computer, influence how you answered the questions?
  1= not at all, 5 = totally
- On a scale from 1 to 5, did you understand immediately how to enter your responses?
  1 = not at all, 5 = totally
- Did filling out the questionnaire influence how you felt? Please specify why.
- On a scale from 1 to 5, how accurately do you think your responses reflect your feelings?
  1= not at all, 5 = very
- On a scale from 1 to 5, how much did the way you were asked about your feelings (i.e. phrasing of the questions) influence the way you answered these questions?
  1 = not at all, 5 = very much
- What is your general feedback about the web questionnaire?

Feedback Questionnaire Smartphone

- On a scale from 1 to 5, how disruptive was it for you to fill out the questionnaire during the day?
  1= not at all, 5 = very
- On a scale from 1 to 5, how much do you feel your environment (e.g. the people around you) influenced the way you answered the questions?
  1= not at all, 5 = totally
- Which time of the day did you find it most intrusive?
  - Early in the morning
  - Late in the morning
  - Right before lunchtime
  - During lunchtime
  - Right after lunchtime
  - Early in the afternoon
  - Late in the afternoon
  - In the evening
- On a scale from 1 to 5, do you think the time necessary to complete a single survey was accurate?
  1= not at all, 5 = very
- On a scale from 1 to 5, do you think filling out the questionnaire 5 times per day on the iPhone was burdensome?
  1= not at all, 5 = totally
- On a scale from 1 to 5, do you think the questions were difficult to understand?
  1 = not at all, 5 = very
- On a scale from 1 to 5, how burdensome was it to fill out the survey?
  1 = not at all, 5 = very
- On a scale from 1 to 5, how easy was it to complete the survey?
  1 = not at all, 5 = very

- On a scale from 1 to 5, did the form of the survey, i.e. that you answered the questions on your iPhone, influence your answers?
  1= not at all, 5 = totally
- On a scale from 1 to 5, did you understand immediately how to enter your responses?
  1 = not at all, 5 = totally
- Did filling out the questionnaire influence how you felt? Please specify why.
- On a scale from 1 to 5, how accurately do you think your responses reflect your feelings?
  1= not at all, 5 = very
- On a scale from 1 to 5, how much did the way you were asked about your feelings (i.e. phrasing of the questions) influence the way you answered these questions?
  1 = not at all, 5 = very much
- What is your general feedback about the iPhone questionnaire?

Feedback Questionnaire Smartwatch

- On a scale from 1 to 5, how disruptive was it for you to fill out the questionnaire during the day?
  1= not at all, 5 = very
- On a scale from 1 to 5, how much do you feel your environment (e.g. the people around you) influenced the way you answered the questions?
  1= not at all, 5 = totally
- Which time of the day did you find it most intrusive?
  - Early in the morning
  - Late in the morning
  - Right before lunchtime
  - During lunchtime
  - Right after lunchtime
  - Early in the afternoon
  - Late in the afternoon
  - In the evening
- On a scale from 1 to 5, do you think the time necessary to complete a single survey was accurate?
  1= not at all, 5 = very
- On a scale from 1 to 5, do you think filling out the questionnaire 5 times per day on the iPhone was burdensome?
  1= not at all, 5 = totally
- On a scale from 1 to 5, do you think the questions were difficult to understand?
  1 = not at all, 5 = very
- On a scale from 1 to 5, how burdensome was it to fill out the survey?
  1 = not at all, 5 = very
- On a scale from 1 to 5, how easy was it to complete the survey?
  1 = not at all, 5 = very
- On a scale from 1 to 5, did the form of the survey, i.e. that you answered the questions on your Apple Watch, influence your answers?
  1= not at all, 5 = totally
- On a scale from 1 to 5, did you understand immediately how to enter your responses?
  1 = not at all, 5 = totally
- Did filling out the questionnaire influence how you felt? Please specify why.

- On a scale from 1 to 5, how accurately do you think your responses reflect your feelings?
  1= not at all, 5 = very
- On a scale from 1 to 5, how much did the way you were asked about your feelings (i.e. phrasing of the questions) influence the way you answered these questions?
  1 = not at all, 5 = very much
- What is your general feedback about the Apple Watch questionnaire?

Overall Feedback Questionnaire

- Overall do you think the time of the day you completed the questionnaire influenced your answers? Please specify.
- On a scale from 1 to 5, how burdensome was it to fill out the surveys
  1 = not at all, 5 = very
- On a scale from 1 to 5, do you think the questions were difficult to understand?
  1 = not at all, 5 = very
- On a scale from 1 to 5, how easy was it to complete the survey?
  1 = not at all, 5 = very
- Do you have any feedback about the design of the questionnaires?
- Was it more natural to complete the survey in the moment or at the end of the day? Please explain why.
- On a scale from 1 to 5, did you understand immediately how to enter your responses?
  1= not at all, 5 = totally
- On a scale from 1 to 5, how accurately do you think your responses reflect your feelings?
  1= not at all, 5 = very
- On a scale from 1 to 5, how much did the way you were asked about your feelings (i.e. phrasing of the questions) influence the way you answered these questions?
  1= not at all, 5 = very much
- If you were sometimes unable to respond, what were the main reasons?
- Comparing the three different surveys, which device type did you prefer? Please rank the three types in the box on the right-hand side. The survey type at the top-most position is the one you preferred the most.
- On a scale from 1 to 5, were these representative 9 days of your life? 1= not at all, 5 = very
- Do you feel that completing the questionnaire in the moment influenced your answers? Please specify why and how.
- Do you feel that completing the questionnaire at the end of the day influenced your answers? Please specify why and how.
- If you have any additional feedback, please leave your comment below.

Fact Sheet

People and Computing Lab
University of Zurich
Department of Informatics
Binzmühlestrasse 14
CH-8050 Zurich

Contact person:
Catrin Loch
Phone +41 76 358 75 78
catrin.loch@uzh.ch

Prof. Dr. Thomas Fritz
fritz@ifi.uzh.ch
University of Zurich

Prof. Dr. Elaine Huang
huang@ifi.uzh.ch
University of Zurich

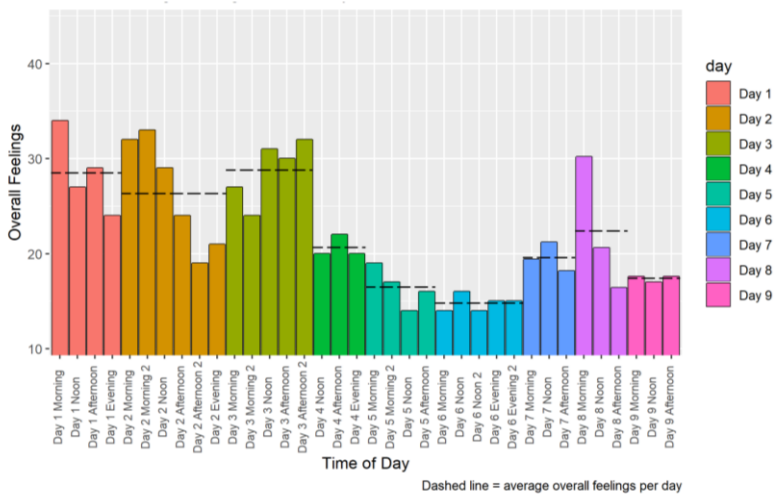
About your Day: exploring methods of self-reporting stress

Dear Claudia

Thank you for participating in this study and taking your time. The aim of this study was to find out what is the "best" method to gather self-reported data about stress and in which way should this data be gathered. To achieve this goal, we investigate different methods of self-reporting activities and feelings and assess people's response to these different methods. You responded to 36 out of 39 assessments throughout the nine days of the study. As a short reminder, you started to fill out the assessments on your Apple Watch, switched then to your iPhone and in the last three days you completed the web surveys.

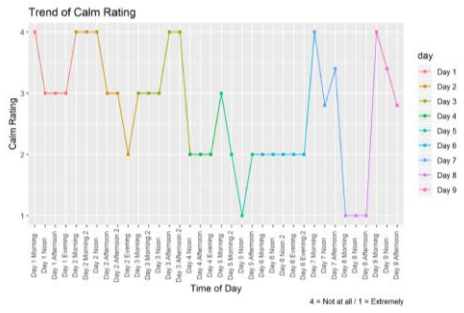
Overall feelings in the course of the study

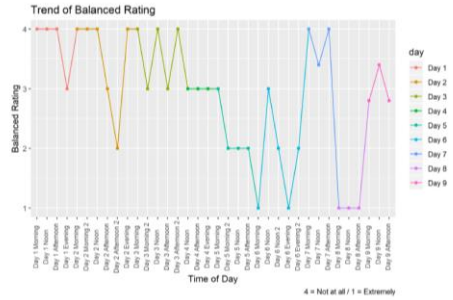
The graph below shows you the aggregated feelings for each assessment you filled out during the study. The overall feeling includes the aggregated value of all eleven feelings, which includes the following feelings: calm, balanced, rested, tense, overcommitted, overwhelmed, stressed, mentally exhausted, in a hurry, under time pressure and anxious. The maximum overall feeling value is 44 and the minimum is 11.

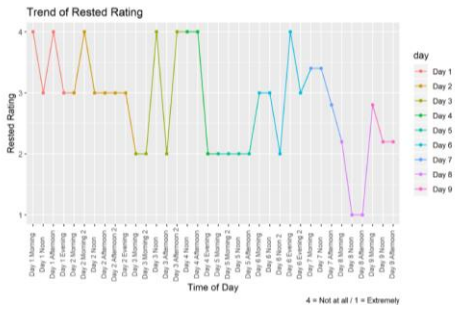


Trend of each feeling

The following graphs show you the trend for each specific feeling for all assessments you filled out. The different colors represent the different study days. The feelings calm, balanced and rested have a maximum value of 4, which means you were not at all calm, balanced or rested. The minimum value is 1, which means you were extremely calm, balanced and rested.







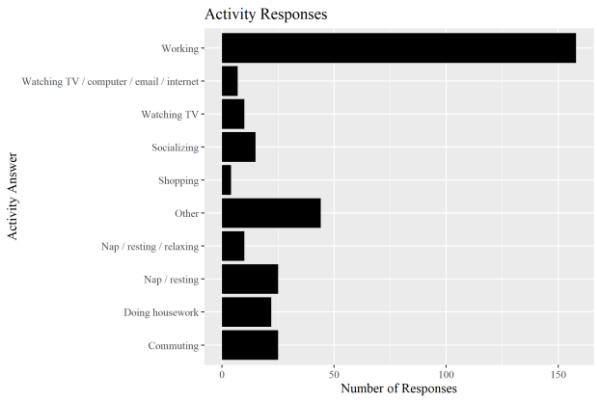
All the following graphs have a maximum value of 4, which means for example you were extremely anxious or mentally exhausted. The minimum value is 1, which means for example you were not at all anxious or mentally exhausted.

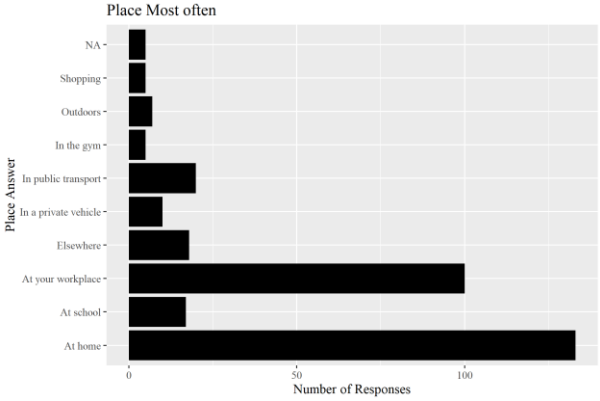
If you have any questions, please feel free to contact me.

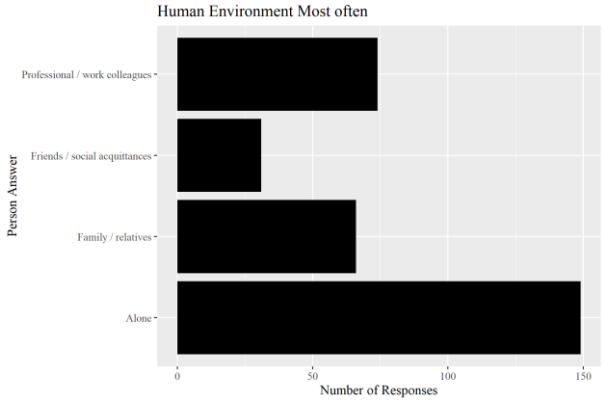
Best wishes,
Catrin Loch

Appendix C: Results

Results Activity, Place and Human Environment







Response Rate

	Participant	Number of Assessments	Response Rate (in %)
1	P01	25	64.10256
2	P02	30	76.92308
3	P03	39	100.00000
4	P04	27	69.23077
5	P05	36	92.30769
6	P06	42	107.69231
7	P07	41	105.12821
8	P08	33	84.61538
9	P09	38	97.43590
10	P10	35	89.74359
11	P11	31	79.48718
12	P12	38	97.43590
13	Total	415	88.67521

Pre-processed Results Feedback Questionnaires

Comfort

Ranking 1

Survey Type							**Count**
Computer	P1	P3	P4	P8			**IIII**
Phone	P7	P11	P12				**III**
Watch	P2	P5	P6	P9	P10		**IIIII**

Ranking 2

Survey Type							**Count**
Computer	P7						**I**
Phone	P6	P8	P10	P9	P2	P5	**IIIIII**
Watch	P3	P4	P11	P12	P1		**IIIII**

Ranking 3

Survey Type								**Count**
Computer	P6	P10	P9	P11	P2	P5	P12	**IIIIIII**
Phone	P3	P4	P1					**III**
Watch	P8	P7						**II**

Overall Questionnaire: Ranking Reason

P6: Convenience. You are not using your phone for everything if you have a watch. For the iPhone survey you sometimes have to take your phone into your hands just to complete the survey - which is burdensome. The Web survey in the end of the day did not feel well, at least for me.

P3: The web survey was easily included in my schedule as it was once per day. I like the iPhone survey the least because when the pop up appeared on my watch, I had to immediately look for the phone. The problem was solved by the watch survey which I put as a second choice.

P8: From a UI/UX perspective I found the iPhone survey the most comfortable one. It was very quick and easy to fill in the survey. However, I preferred the web survey a bit more simply because it was just one survey a day and it was in the evening when I usually had time to fill it in. It felt less interrupting than the iPhone surveys.

P7: iPhone was the most natural way to answer as it was quickly available and convenient to answer. The web survey was convenient as well but I think my answers don't truly reflect my feelings at a certain point since it was hard to think back how I felt at a certain point. The apple watch survey was slightly annoying as I had to keep it on my wrist all day long and answering questions on such a small screen was sometimes annoying (especially at the gym or in public transport)

P10: Filling out the survey on the watch is less interruptive than on the phone. Especially, when I was with a group of people. Filling it out on the watch and explaining what it was, was interesting to my colleagues. However, getting the phone out seemed more like I don't want to interact with them and it was considered ruder.

P4: I like to think about my answer (and change possible inputs). But I guess the questionnaire is intended to capture emotions that are more automatic and not deliberative.

P9: The watch was sooo much faster for me. I did not need to pull out my phone or really interrupt what I was doing. The phone was similar but not quite as convenient, as I do not always have my phone with me during the day. The web survey was the most annoying for me and felt like it takes the longest (even if that probably is not true)

P11: iPhone: good in accuracy and receiving notification. Apple Watch: good in accuracy, bad in receiving notification. Web survey: bad in accuracy (recalling emotion is not easy)

P2: iWatch was the best since it is very subtle and i had finally something to really do with the watch where it is usally just showing information where the full story has to be read on the phone. The phone was also ok. but less handy than the iwatch, i also prefered that there was a back button in the watch app which did not exist in the iphone app. I really didnt like the Web app since there every thing seemed so distant and for me püroblematic

was that the previously awnsered questions were not visible but for me it would have been much easier since usually i rememberd when i had a good time and a bad time like a mood curve but without knowing what i previousl answered it seemed a bit random for me.

P5: I did not really like the web survey and would not use it in my routine. I could imagine to use the watch survey. My main problem with the web survey was, that it is a reflection over the entire day and not on the current situation. I liked the watch survey better than the phone survey because I immediately felt the notification and could answer it without searching the phone in the bag and could easily and fast answer the questions. I was happy about the watch notifications when using the phone survey, but if this would not have been notified on the watch as well, I might not see the notification on the phone. Because I usually have my phone in night shift mode which mutes all notifications and turns off vibrations.

P12: All the answer possibilities could be better seen on the phone. the display of the watch is a bit too small for that, however, on the watch through the vibration I realized it more soon when e new survey was available

P1: I could take my time to fill out the web survey in the evening and think my answers were more accurate then. Since I always have my watch with me, I could do the Apple Watch survey immediately and it was easy to handle. I missed some notifications on the phone easily during the working days.

Computer Questionnaire: General Feedback

P7: It was interesting but I regret that the time scope did not go through the evening as I am convinced that feelings in the evening as really important as well. Maybe include "studying" when asked "what were you doing at that time" there is mention for that.

P10: I find the 6-point Likert scale (if I remember correctly) difficult to answer. For me, I am either stressed, not stressed or in the transition. This transition, however, is very difficult to judge for me. I wish it was more like a 3-point or 4-point scale rather than 6-point.

P1: It is easy to understand the way it works and didn't take a lot of time. I found it difficult to scale my feelings and I'm not sure how accurate it would be.

P4: It was not possible to specify different types of people I met during the 3 hours (3 hours are quite a long time)

P9: It just seemed too take longer, even-though objectively it probably did not take as much time as the phone surveys of a day combined.

P12: The questionnaire is self-explenary, easy to understand and easy to fill in.

P6: As said before: Answering these questions in retrospective is an estimate and mobile/app are probably more accurate. Usability of the web survey is good, but having to fill it out every evening - where also other things have to be done - is a little tiresome.

P3: The web survey was my favorite one as I could easily include it into my daily schedule. However, in comparison to the watch and iPhone surveys, the results could not portray my emotions so specifically as I sometimes had difficulty recalling feeling particular emotion or state.

P8: I found this one probably the least intrusive one. First, because it was only one survey a day and second, because I usually had time in the evening to fill in the survey and didn't feel interrupted by it as much as the Watch/iPhone surveys.

P5: I would not use the website to answer questions about my day. Because when I am away from work, I usually do not use a laptop/computer and mainly that's why I would not use it. Moreover, I did not really like to reflect every day at the end of the day over the entire day and put each situation and feeling in perspective (sich noch mals in die Situation versetzen).

P11: Recalling my feeling during the day was way more difficult than I thought.

Smartphone Questionnaire: Overall Feedback

P9: Doesn't always vibrate when a new survey is there, so I missed it sometimes.

P3: I believe the questionnaires had a positive effect on me. After answering some of the questions, I took my time to think about particular emotions being connected to the place I am right now, the people (or lack of them) around me as well as the activity I was performing. Filling the survey on my iPhone was always quick.

Notifications were a nice way of reminding me about the survey as keeping track with a paper schedule would be problematic.

P4: I think the questionnaire was easy to understand and fill in. However, I missed the back button. Moreover, I encountered 2 unintended exits from the app (Absturz)

P11: It was okay in general. Maybe it's because of the app, but sometimes I received multiple notification together and what I did was to answer them one after the other in ten minutes or so. I am not so sure if that's the expected behavior.

P7: I think you should have asked questions about how I felt during the evening as well. Otherwise, it was really easy to follow the questions and respond in time.

P5: Unfortunately, I had the situation that I was interrupted during the survey by another person, and then my phone went into the lock mode and when I wanted to continue the survey, it was already submitted and I could not complete it entirely. Before I noticed this, it already happened a few times, I am sorry for that. Overall, I did not really like to fill out the questionnaire on the phone. I guess it is mainly because I do not really want to be on the phone during work. Moreover, I usually have the phone on the 'night mode' which turns off all vibrations, in order to not disturb my colleagues. Therefore, I really liked that the notifications also popped up on the watch and I saw that I had to fill out a new questionnaire.

P1: It was easy to handle. Nevertheless, I think 5 times a day are too much.

P8: I found the questionnaires on the iPhone more user-friendly than on the Watch, simply because the interface was easier to use with the bigger screen and the better responsiveness of the app. I did miss more questions in the iPhone session because I was out with friends a lot during this time and didn't see the notifications or didn't want to interrupt the interactions with my friends.

P6: The app was easy to use, no errors or whatsoever. Certainly a good alternative to the watch version.

P10: The iPhone questionnaire was much much nicer compared to the web version. Filling it out on the phone is a lot quicker. However, when I was in a conversation it seemed to be rude to answer the questionnaire on the phone. With the watch this seemed to be less of an issue.

P12: It is really easy to fill in and only takes a few seconds. If you press the wrong button, you cannot go back, that is sometimes a challenge, but it is better as you cannot take too much time to correct your answers.

P2: The iPhone part should have a back button. When the app is closed during the survey it can't be finished.

Smartwatch Questionnaire: General Feedback

P6: Interesting, but a little too often in my sense. The menu was easy to use (on a 44mm apple watch), so: Well done! Sometimes I did not get notifications but realized a while later that there has been a survey available, just fyi.

P5: I was not sure if this functionality exists, but I had the impression not: If one has no internet connection (or another failure to send the survey), the survey should be stored locally and send as soon as the internet connection is available. Otherwise, I have to remember to open the App again and submit the survey.

P8: I think it worked quite well, since the interaction with the watch is simple and short. I would probably still prefer filling in the survey on the iPhone because the UI is bigger and my iPhone is faster.

P12: It was really easy to answer, short questions, only a few questions. Sometimes the system was not answering immediately so I pressed submitted more than once. In general, it was not time consuming.

P11: It was a bit tedious because I didn't always receive the notification via the watch. Often I see several notifications on my iPhone once together.

P3: I preferred the Apple Watch surveys over the IPhone ones as they were easier to fill. Overall, I did not notice a change in my emotional state influenced by a particular device.

P10: Generally speaking, it was very lightweight and easy to do. However, I made some mistakes. For example, "feeling anxious" --> "not at all" which is on the left. "calm" --> "extremely" on the right. I sometimes accidentally pressed the button on the wrong side. There was no possibility to correct this error.

P2: The scale from not at all to extremly was a bit strange, positive feelings sound strange together with the word extremely 'extremely calm' sounds some what negative for me, like a little too calm.

P7: This was the less effective questionnaire of the three (computer, iPhone, watch) because I didn't necessarily had my watch on my wrist all day long and sometimes the notification appeared on my phone but not on the watch so I had manually open it in the watch. Also filling a questionnaire on such a small screen is not really convenient. However, I think this could be useful to use this apple watch to measure heartbeat rate at the time of the questionnaire (which we can't do on the iPhone)

P9: I think once the technical issues (vibration and that it pops up while the hr is being recorded) are solved, it will be the best version to use. At least for me, as I don't like having my phone around the whole time, so being able to do everything on the watch is great.

P1: Easy to handle, less disruptive than other questionnaires.

P4: It was better as expected. Sometimes it did not send my inputs or I was not sure whether it did. Despite that it was very easy to fill in and captured my emotions of the moment.

Computer Questionnaire: Usability

Participant	Value	Participant	Value
P01	5	P07	5
P02	5	P08	5
P03	5	P09	5
P04	4	P10	5
P05	5	P11	2
P06	5	P12	5
Total	56		
Average	4,7		
Coding	5		

P6: I think it was pretty straight forward.

Smartphone Questionnaire: Usability

Participant	Value	Participant	Value
P01	5	P07	5
P02	5	P08	3
P03	5	P09	5
P04	5	P10	4
P05	5	P11	4
P06	5	P12	5
Total	56		
Average	4,7		
Coding	5		

P3: The description, schedule as well as the survey content was well-structured and easy to be understood.

P4: was very clear and easy to use

P8: The same problem as on the Apple Watch, I sometimes had to think for a second which side of the scale was "positive" and which side was "negative" answers. E.g. For "Do you feel overwhelmed?" the "positive" values were the low ones (1, 2) whereas for "Do you feel calm?" they were the high values (3, 4).

P6: No problem.

P2: Very easy and accessible

Smartwatch Questionnaire: Usability

Participant	Value	Participant	Value
P01	4	P07	5
P02	5	P08	4
P03	5	P09	5
P04	5	P10	5
P05	5	P11	3
P06	5	P12	5
Total	56		
Average	4,7		
Coding	5		

P8: For some questions the my intuition of what 1 represents and what 4 changed. E.g. I found myself pressing 1 or 2 a lot for "negative emotions" ("do you feel anxious / stressed / overwhelmed") and then had to make sure I didn't press 1 or 2 for "positive emotions" ("do you feel balanced / rested ") even though I felt balanced / rested and should have pressed 3 or 4. I connoted 1 for "good" and 4 for "bad" automatically. But depending on the question 1 might be "bad" and 4 might be "good".

P3: Everything was precisely described in the schedule.

P9: If you've ever used the watch then its simple inputs like everywhere else.

Overall Questionnaire: Usability

Participant	Value	Participant	Value
P01	5	P07	5
P02	5	P08	5
P03	5	P09	5
P04	5	P10	5
P05	5	P11	3
P06	5	P12	5
Total	58		
Average	4,8		
Coding	5		

P6: No problem at all.

Overall Questionnaire: Naturalness

P6: As explained, I liked the ones in the moment better.

P3: In the moment. Recalling particular emotions from the morning hours while filling the survey in the evening was sometimes challenging.

P8: I would say recording my mood in the moment several times throughout the day gives me a better insight into my emotions, because emotions can change quite drastically in the course of a day. Also, when filling in the survey in the evening/end of the day I might not remember exactly how I felt in the morning. Also the questions in the end-of-day survey spanned over three hours, which made giving answers more difficult because I might have felt very calm at the beginning of those three hours but felt stressed at the end of them.

P7: In the moment because it was way easier to truly report my feelings.

P10: I prefer giving it right at the moment. This way the memory is still fresh, and most of the times it was not really interruptive. However, when I was sitting in a meeting, I filled it out right after.

P4: I would not say natural, but perhaps more spontaneous and more honest - however I did sometimes mistakes on the phone or watch and could not go back.

P9: Definitely in the moment! I was able to just put in how i was currently feeling and did not have to think back and try to remember how I was feeling.

P11: The moment, because of the accuracy of emotion

P2: It is much betetr to do it in th emoment, since one has to think about the day and decide how one felt at a specific point in time, but for me most of the time this was very difficult or i just remembered a vague mood curve.

P5: Definitely in the moment. If I had to think about the day and rate the feelings from the morning, I felt that my answer might have been influenced by the feelings which happened afterwards (e.g. the afternoon). Moreover, for me it was hard to look at the feelings and rate them depending on their time-occurrence. e.g. just think about the feelings in the morning. Except of the feeling which were very intense and remarkable, but not a usual feeling.

P12: As it pops up on the screens like massages etc. I think we are conditioned to answer it immediately. So I found it natural to answer it immediately, like a task, hence, I had one thing done from my „to do list".

P1: It was more natural completing the survey at the end of the day. I could take my time to focus on the survey and reflect my thoughts.

Overall Questionnaire: Design

P6: Sometimes it was quite hard to assess the right answer as it's quite a jump between "moderate" and "not at all" (or the other). I'm quite sure that you covered the why in the design of your questionnaire, but I wonder if an additional answer (or two) wouldn't help.

P3: It was simple and straightforward which made filling the survey quick. Various pop ups served as a great reminder. The only thing changing were the adjectives which saved time on reading the whole sentences all over.

P8: It so happens that I was eating a few times when filling in the survey, but eating wasn't one of the answers in the activities question. I selected "other" instead.

P7: In general, it was really good but I was expecting a more in-depth feelings analysis (more questions about other sentiment traits)

P10: I gave some feedback above.

P4: Use the same scale across input types
extremely calm sounds akward to me

P9: I think it was a good amount of questions. However, I did think that filling out the same questions for different times of the day on the desktop survey was a bit boring.

P11: The scale is a bit vague

P2: My feeling is that most of the feelings asked are more negative than positive. The four choice questionnairs were much easier to answer then the one with 7 possibilities. The Web questionnair was most difficutl and annoying to fill in if all these questionnairs are comapred. I Prefered the iwatch sicne it is the most subtle way to fill in the questionnair and it is very handy since no diveice has to be picked out of the pocket it can just be done, when ever needed.

P5: Yes. In the question 'Were you with anyone' I missed the answer option 'with random people', because sometimes I was surrounded or had to deal with persons which were neither my colleagues, friends or family.

P12: It would be helpful to go back if one pressed something wrong, but on the other side it is good that one cannot go back to complete the survey in a reasonable time.
Otherwise I find the design simple to understand and easy to handle and clear.

P1: Some adjectives are pretty the same (e.g. calm and rested) and it was difficult to differentiate.

Intrusiveness

Participant	Value	Participant	Value
P01	1	P07	1
P02	1	P08	2
P03	1	P09	5
P04	1	P10	2
P05	2	P11	1
P06	4	P12	2
Total	23		
Average	1.9		
Coding	2		

Computer Questionnaire: Disruptiveness:

P10: Sometimes, the saw the e-mail before I wanted to leave. I stayed longer to fill out the form as I didn't have plans in the evening. Maybe, if participants would know when the questionnaire is sent (e.g. 17:22), we could plan a bit better.

P4: Was at a time of the day when I was calm

P9: Needing to go get my laptop was a hassle.

P12: Only when I was still at work it was a bit disruptive.

P6: In the end of the day you're normally resting and filling out the survey kind of led you back to "work". Filling out the app-survey disturbed me less, I feel.

P3: It was more convenient to fill one survey than a few during the daytime.

P8: I was usually at home in the evening and had time to fill in the survey.

Participant	Value	Participant	Value
P01	4	P07	2
P02	2	P08	3
P03	2	P09	2
P04	3	P10	4
P05	5	P11	4
P06	4	P12	2
Total	37		
Average	3,1		
Coding	3		

Smartphone Questionnaire: Disruptiveness

P9: Filling out the survey wasn't as much of an issue, as stopping what you were doing and then restarting after.

P3: Sometimes the survey appeared while I was having a Skype meeting with my employer. As it was required for me to fill the survey as soon as possible, I had to keep it in mind for the whole meeting.

P4: There were times at which I was not able to take my phone out of the bag

P1: I have not received any notifications on my phone on the first day of phone survey. Furthermore, it was sometimes untimely for filling out the questionnaires.

P8: It was sometimes disruptive on the weekend because I was out with friends and didn't check my phone as often. Or sometimes, I didn't want to interrupt a conversation with a friend to fill in the survey.

P6: I rated it 4 in comparison to web- or watch-based. You're not using your phone so much anymore if you have an apple watch, so it felt like I have to pick up my phone "just for the survey". If someone has no watch, I would prefer the app over the web-based survey.

P10: It was a little more disruptive than the watch but less compared to the web version.

P12: It is done really fast, so it is not really disruptiv

P2: Depending on the task at hand but usually not too much.

Participant	Value	Participant	Value
P01	4	P07	2
P02	2	P08	4
P03	3	P09	1
P04	3	P10	2
P05	2	P11	2
P06	3	P12	1
Total	29		
Average	2,4		
Coding	2		

Smartwatch Questionnaire: Disruptiveness

P6: In between of work its sometimes difficult to fill out the survey without losing the line of thought.

P5: When I was talking to someone, I felt it more distruptive and I usually postponed the answer, which led to the problem, that I forgot it later.

P8: It often interrupted my workflow.

P3: As the pop ups appeared on my watch, I had to wear it during the weekend when I normally don't.

P10: It is super quick to do on the watch.

P2: It was very easy to fill in the survey while doing my job, but more difficult when 100% of my attention was needed during a longer period.

P9: It was super quick, so it didn't really bother me.

P4: As the watch vibrated softly, It was not that disruptive.

Computer Questionnaire: Environment Influence

Participant	Value	Participant	Value
P01	4	P07	5
P02	1	P08	1
P03	2	P09	3
P04	1	P10	3
P05	1	P11	2
P06	1	P12	3
Total	27		
Average	2,3		
Coding	2		

P1: I might tend to answer more positively.

P4: Was always on my own.

P9: Needing to interrupt my evening, made me more annoyed, which might have influenced my answers a little.

P12: It depends if they influence the feelings, if they are a part of / or the cause of the feelings.

P6: Not at all. I would not know how or why.

P5: I was not surrounded by anyone when I filled out the survey.

Smartphone Questionnaire: Environment Influence

Participant	Value	Participant	Value
P01	1	P07	4
P02	1	P08	2
P03	1	P09	2
P04	1	P10	3
P05	1	P11	5
P06	1	P12	3
Total	25		
Average	2,1		
Coding	2		

P9:Some situations hindered me from answering (like meetings)

P3: For all the surveys I was alone. However, I completed the surveys faster while being at work than, e.g. while commuting.

P6: I don't know why they would.

P2: I don't think that there is an influence between surrounding people and answering a questionnaire.

Smartwatch Questionnaire: Environment Influence

Participant	Value	Participant	Value
P01	2	P07	5
P02	1	P08	2
P03	2	P09	2
P04	2	P10	3
P05	1	P11	3
P06	1	P12	3
Total	27		
Average	2,3		
Coding	2		

P6: Not that I realized - but I worked mostly alone throughout the week so far.

P5: They haven't seen what I am answering, otherwise, it might have had an influence.

P3: It definitely had influence on my emotions but not on a way I fill out the survey.

Overall Questionnaire: Timing of self-reports

P6: I'm not sure regarding the question. If it's about if the time of day would have made me give unprecise answers, then: No, I don't think so. If it's about if I generally saw that at given times of days I was more balanced/frustrated/stressed etc or if I felt more disturbed when filling out the survey, then: Yes.

P3: Yes, I believe it had an impact, as for example, during the morning hours I was more eager to mark higher for positive emotions.

P8: I don't think so. Maybe, I felt more stressed in the afternoon when my work day came to a close and I wasn't done with my work for today yet. So in the afternoon surveys I might have felt more stressed than in the morning.

P7: I think that when answering questions at the end of the day there is probably a bias due to the fact that it is hard to remember how I felt at a certain point in the day. However, when answering questions on the iPhone or apple watch it didn't influence my answers as I was able to answer straight away how I felt at that time.

P10: Yes, I think so. In the morning, I feel more mentally fit, rested, and balanced. Throughout the day these qualities decrease due to the work I do.

P4: I think that it is likely that I filled the questionnaires differently compared to earlier times in the day.

P9: I was more annoyed filling out the surveys in the late evenings, so the computer surveys were maybe influenced a little bit

P11: Hard to say, but the email one (recalling my feeling at three different time on the same day) is surely very difficult

P2: Wheter one is stressed or not does probably change the answers since less time can be spent on answering or reflecting on the questions sow i would guess that answers are less accurate in the after noon when one wants to go home from work and relaxe.

P5: More in a sense of experiencing different feelings at different times of the day, but not that I rated the survey differently at different times e.g. more positively because it was morning.

P12: In one sense, yes, because your feelings depend on the things happening and people being around you, so the more time of a day passes the more could have happened to influence the feelings. However, in general I cannot say the later I completed the questionnaire the worse were my answers. It could have been in both ways negative-> positive and visa versa.

P1: I don't think so.

Overall Questionnaire: Influence Moment

P6: I wouldn't say so. I tried to answer the questions independent from the modality.

P3: I do not think so, however it was much easier to answer as it did not require recalling anything.

P8: No, I could give honest answers whenever I filled in the survey. However, I missed a few surveys because I didn't have time or didn't see the notification on my phone right away in that moment.

P7: No, I don't think so. I had no incentives to misreport my true feelings at time T.

P10: I already answered this in an earlier question.

P4: More erroneous and more spontaneous than in the evening. As I did the questionnaires 'on-the-go', I pressed the wrong button and could not go back. But in all, I did not think long when pressing the button on the phone and watch. I expect the inputs more volatile (or higher in variation)

P9: I think it was more accurate.

P11: Not really

P2: The answers are more accurate.

P5: No. I did not felt this.

P12: yes it depended on the situation and the influencing factors like work load, deadlines, private things to take care of

P1: No. I would come up with the same answers later.

Overall Questionnaire: Influence End of the Day

P6: I don't think so. Even though I liked the web-based surveys the least, I tried to give accurate answers.

P3: yes, as I was very often tired which could result in marking negative emotions higher, even for the morning hours.

P8: I don't think so. Sometimes I didn't remember exactly how I felt in the morning maybe.

P7: Probably as once the day has passed it's sometimes difficult to think back about my true feelings. Or sometimes the day could have started in a very negative way but then end in a positive way and in the end when asked how I felt in the morning I would report weaker negative feelings than it truly was.

P10: I already answered this in an earlier question.

P4: The answers were more in the middle as I think that I thought more about my answers (and changed them during the filling in). I expect less variation when filling out the web survey.

P9: It was sometimes hard to remember how I was feeling throughout the day.

P11: Yes, because I cannot recall my exact emotion

P2: The anwsers are less accurate. Thinking about parts of the day which i only vaguely remember made me a bit aggressive, thats why i really disliked the web questionnaire since i had the feeling that my aswers are not accurate at all and that filling in the form is a bit a waste of time.

P5: Yes, I think they were less accurate because it is more difficult to think back to a certain time and think about my feelings.

P12: Yes, because I could distance myself more from what was happening all day long. Often in the evening I could answer in calm situation as I was on my way home or at home

P1: No. I would give the same answers earlier as in the evening.

Smartphone Questionnaire: Time of day

	P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11**	**P12**	**N**	**Y**
Early in the morning	Yes	No	Yes	No	No	No	Yes	No	No	No	No	No	9	3
Late in the morning	Yes	No	Yes	No	No	Yes	No	Yes	Yes	No	Yes	No	6	6
Right before lunchtime	Yes	No	Yes	Yes	Yes	No	No	Yes	No	No	No	No	7	5
During lunchtime	No	No	No	No	Yes	No	No	Yes	No	Yes	No	No	9	3
Right after lunchtime	Yes	No	No	No	Yes	Yes	No	Yes	No	No	No	No	8	4
Early in the afternoon	Yes	Yes	No	Yes	No	No	No	Yes	No	No	No	No	8	4
Late in the afternoon	Yes	No	No	No	No	No	No	Yes	Yes	No	No	Yes	8	4
In the evening	No	No	No	No	No	Yes	No	Yes	No	No	Yes	No	9	3

Smartwatch Questionnaire: Time of day

	P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**P10**	**P11**	**P12**	**N**	**Y**
Early in the morning	Yes	No	Yes	No	No	No	Yes	No	No	Yes	No	No	8	4
Late in the morning	No	No	Yes	No	No	Yes	No	No	No	No	No	Yes	9	3
Right before lunchtime	No	No	Yes	No	No	No	No	Yes	No	No	No	No	10	2
During lunchtime	No	No	No	No	Yes	No	No	No	Yes	No	No	No	10	2
Right after lunchtime	No	No	No	No	No	Yes	No	No	No	No	No	No	11	1
Early in the afternoon	No	Yes	No	No	No	Yes	No	No	No	No	No	No	10	2
Late in the afternoon	No	Yes	No	No	No	No	No	Yes	No	No	Yes	Yes	8	4
In the evening	No	No	No	Yes	No	Yes	No	No	No	Yes	No	No	9	3

Participant	Value	Participant	Value
P01	4	P07	2
P02	1	P08	5
P03	4	P09	4
P04	3	P10	3
P05	1	P11	2
P06	3	P12	5
Total	37		
Average	3,1		
Coding	3		

Computer Questionnaire: Accuracy of duration of questionnaire

P7: It was closer to 15 minutes than the 30 minutes specified

P4: Could imagine that the outcomes of the day influenced my overall impression of my mood

P12: Even when entering specifying answers, it never took 30 minutes.

P6: I think your estimations have been a little high (also this review, I guess, will not take 40 minutes of every survey participant).
Time-wise, I didn't feel the web-survey to take too much time. It was quite to the point.

P3: The survey did not require much time. I believe the 30-min estimation was too much as you could feel the survey in approximately 15 minutes due to its simplicity.

P5: It took me never 40 minutes. I was a little bit surprised the first time, because I thought I might have forgotten to answer a question or not all questions were published.

P2: Took much less then 40 minutes I think it took about 5 but never stopped the time.

Participant	Value	Participant	Value
P01	4	P07	5
P02	4	P08	5
P03	4	P09	5
P04	4	P10	5
P05	5	P11	5
P06	1	P12	5
Total	52		
Average	4,3		
Coding	5		

Smartphone Questionnaire: Accuracy of duration of questionnaire
P9: It was super quick!

P6: Provided estimations were too high, in my opinion. I was able to complete the surveys faster.

Smartwatch Questionnaire: Accuracy of duration of questionnaire
P6: Not too long, not too short.

P12: Short and simple questions

P3: The number of questions as well as simplicity of answers made the survey quick.

P2: Was rather short to fill out the survey.

P9: It was very fast

P4: Was very short to capture feelings. And if yes only of the time I filled the information in

Participant	Value	Participant	Value
P01	4	P07	5
P02	1	P08	4
P03	5	P09	5
P04	2	P10	4
P05	5	P11	3
P06	4	P12	5
Total	47		
Average	3.9		
Coding	4		

Computer Questionnaire: Burden

Participant	Value	Participant	Value
P01	1	P07	1
P02	4	P08	2
P03	1	P09	4
P04	2	P10	1
P05	2	P11	3
P06	2	P12	1
Total	24		
Average	2,0		
Coding	2		

P10: It was a fairly short questionnaire and at the end of the day, it is fine filling it out. There was not a lot of cognitive load required to do so.

P9: Getting my laptop out of my bag in the evening, going to sit down to fill out the survey and then putting the laptop away again, was pretty annoying.

P12: it takes longer than answering by apple watch, however, it also asks for optional clarifications of the answers

P6: effort-wise not a big problem - just the "going-back-to-work"-aspect is a little bit burdensome.

P3: As I knew when approximately I should fill out the survey, it was included into my daily schedule.

P8: The surveys were longer than the Watch and iPhone surveys. But filling them in on the PC felt very easy.

P2: I was really annoyed by the web questionnaire since it is hard for me to remember short term stuff.

Smartphone Questionnaire: Burden

Participant	Value	Participant	Value
P01	5	P07	2
P02	1	P08	4
P03	4	P09	3
P04	5	P10	2
P05	4	P11	3
P06	3	P12	1
Total	37		
Average	3,1		
Coding	3		

P3: As I did not know how many times it would be requested for me to fill the survey, I felt pressured to check my phone more frequently than usual.

P4: I use my phone not that much except for communication.

P8: I missed quite a few surveys because I was busy with friends on the weekend and didn't check my phone often.

P6: As explained above, It's ok on a phone if you don't have a watch - otherwise I would prefer the phone. Overall, it wasn't too disturbing - I would rate it 3.

P2: Again this is very depending on the tasks thet had to be fulfilled. On day I had a meeting the whole day which made it difficult to fullfill. iPhone and iWatch are much better then once on the computer. The usability of the watch was the best even though it is unclear wether the finger always hit the right number.

Smartwatch Questionnaire: Burden

Participant	Value	Participant	Value
P01	2	P07	1
P02	1	P08	4
P03	2	P09	2
P04	1	P10	1
P05	1	P11	4
P06	3	P12	1
Total	23		
Average	1,9		
Coding	2		

P8: It seemed like a lot of reporting. It's alright for a study like this, but I don't think I would have the discipline/will to keep this frequency of reporting up for a long time.

P12: it takes only a few seconds to fill in the questionnaire

P11: I don’t always get notification via my Apple Watch.

P3: Even though the pop ups served as a great reminder, I subconsciously kept in mind the requirement of filling further surveys.

P2: Very comfortable, but seems not very accurate or hard do know what was actually answered.

P9: i enjoyed it a lot more than needing to take out my phone, as i was able to fill it out as soon as the notifications came (if i noticed it, as it did not vibrate)

P4: Very short and rhus very small amount of time

Overall Questionnaire: Burden

Participant	Value	Participant	Value
P01	2	P07	1
P02	2	P08	4
P03	3	P09	2
P04	3	P10	2
P05	2	P11	3
P06	2	P12	1
Total	27		
Average	2,3		
Coding	2		

P6: As specified before, the daily web-based surveys were the most burdensome, while the others were manageable. At the beginning filling out 5 surveys a day was a little burdensome, but I got used to it.

P3: The watch and iPhone surveys, even though they were quick and simple, required constant checking for pop ups.

P8: I found the interruption of whatever I was doing irritating in the Apple Watch and iPhone surveys, especially because there were quite a few surveys a day. And it does take some time off my day to fill in the surveys.

P10: Filling out the survey was fairly simple. The number of questions was appropriate to not interrupt the task/activity too much.

P4: It was complicated to get out the phone during the days just for the questionnaire. I preferred the questionnaires in the evening (web feedback) despite they might be biased

P9: It did not take a lot of time, especially the surveys on the watch are super fast.

P12: Simple questions, which in my opinion should be answered spontaneously (Bauchgefühl), to lead to the most propriate answers. Short questionnaire

Smartphone Questionnaire: Burden 5 Times

Participant	Value	Participant	Value
P01	2	P07	1
P02	1	P08	1
P03	2	P09	2
P04	1	P10	1
P05	4	P11	3
P06	3	P12	1
Total	22		
Average	1,8		
Coding	2		

P8: I found filling in the survey on the phone less cumbersome than on the Apple Watch because the interface was faster and the buttons bigger.

Smartwatch Questionnaire: Burden 5 Times

Participant	Value	Participant	Value
P01	2	P07	2
P02	1	P08	2
P03	3	P09	1
P04	1	P10	1
P05	1	P11	2
P06	3	P12	1
Total	40		
Average	1,7		
Coding	3		

P8: I find longer interactions on the Apple Watch cumbersome, just because of the tiny interface and because my Watch (Series 2) is quite slow. For this quick survey I thought it worked quite well though as the interactions were not too long (it usually didn't take me more than 1 min to fill in the survey) and the UI was responsive.

P3: Sometimes it made me think of some of the emotions and a possible reason for them which disrupted my work flow.

P2: was very easy and did not take much time

P9: Clicking the buttons was super easy, the need for scrolling was less great but totally doable.

Computer Questionnaire: Understanding of Questions

Participant	Value	Participant	Value
P01	1	P07	1
P02	1	P08	1
P03	1	P09	1
P04	1	P10	3
P05	2	P11	4
P06	1	P12	1
Total	18		
Average	1,5		
Coding	1		

P10: Some questions were phrased very similarly. I had to read some of them multiple times to understand the difference.

P6: I felt they have been quite to the point - no problem here.

P8: Somehow, I didn't encounter the problem which I had in the Watch and iPhone surveys where I had to think about which extreme (positive vs. negative) is on which side of the scale. Maybe it has something to do with the UI. On the Watch/iPhone the next question appeared in the exact same spot as the previous one, so I didn't have to move my finger to select the same answer (e.g. a 1). On the web, I had to move the mouse either way, maybe that reduced my bias to always think of "left" as the "positive" extreme and "right" as the "negative" extreme.

Smartphone Questionnaire: Understanding of Questions

Participant	Value	Participant	Value
P01	1	P07	1
P02	1	P08	2
P03	1	P09	3
P04	1	P10	1
P05	2	P11	4
P06	1	P12	2
Total	20		
Average	1,7		
Coding	1		

P9: Many of the feelings were opposites, if it asks me if i am anxious and i say very then it doesn't need to ask me if i'm calm.

P6: They were straight forward to me.

P12: Some of the different feelings are pretty similar so it is hard to distinguish

P2: Same as in all the other parts of the study.

Smartwatch Questionnaire: Understanding of Questions

Participant	Value	Participant	Value
P01	1	P07	1
P02	2	P08	2
P03	1	P09	1
P04	1	P10	1
P05	2	P11	2
P06	1	P12	2
Total	17		
Average	1,4		
Coding	1		

P8: Maybe the "Do you feel overcommited" question wasn't quite clear to me at first. I interpreted it as having too many commitments right now (more than I could handle).

P12: some of the questions were pretty similar, whats makes it difficult to distinguish

P2: Well some feelings were overlapping like stressed and in a hurry which might be confusing.

Overall Questionnaire: Understanding of Questions

Participant	Value	Participant	Value
P01	1	P07	1
P02	2	P08	1
P03	1	P09	1
P04	2	P10	4
P05	2	P11	2
P06	1	P12	2
Total	20		
Average	1,7		
Coding	1		

P6: I don't think so, they were straight forward to me.

P3: Phrasing was simple and requirements straightforward.

P10:Especially when I was mentally exhausted, I had to read the questions multiple times to see how they differ. Each question as posed very similar.For example, most of the questions start with "On a scale from 1 to 5, ...", which is implicitly clear (at least for me). I wish the questions would be phrased something like "How easy was it to complete the survey? \n 1 = not at all, 5 = very.

P4: I think the time intervals, e.g. 8to 11 am, were too long to just insert one category of people I interacted with.

P9: Really simple and straightforward questions

P12: Because the feelings are similar

Computer Questionnaire: Easiness

Participant	Value	Participant	Value
P01	4	P07	5
P02	3	P08	5
P03	2	P09	5
P04	5	P10	5
P05	5	P11	3
P06	4	P12	5
Total	51		
Average	4,3		
Coding	5		

P6: Overall very good. But you could consider having three columns with each of the three times of the day you were asking about next to each other - this would allow for a faster filling out of the survey. Also, people might be more accurate when they can compare directly how their mood changed over their day.

P3: Sometimes I had difficulty recalling particular emotions during the morning hours.

P2: Not so easy it was hard to decide between the 7 possibilities.

Smartphone Questionnaire: Easiness

Participant	Value	Participant	Value
P01	4	P07	5
P02	5	P08	5
P03	4	P09	5
P04	5	P10	5
P05	4	P11	4
P06	5	P12	5
Total	56		
Average	4,7		
Coding	5		

P3: The questions were easy to be understood and the answers quite general.

P4: It was easy to click the structured questions. Some times, questions were akward , e.g. "feeling extremely rested"

P5: Good usability,

P8: Very simple process. The only thing missing on the phone was to go back to a previous answer (which I think was possible on the Apple Watch?).

P6: I did not see a major difficulty - questions and structure seemed very well prepared.

P12: In general, very easy. However, once I left the app before submitting and then you cannot reopen the survey on your one.

Smartwatch Questionnaire: Easiness

Participant	Value	Participant	Value
P01	2	P07	5
P02	5	P08	4
P03	3	P09	4
P04	5	P10	5
P05	5	P11	4
P06	5	P12	5
Total	52		
Average	4,3		
Coding	5		

P6: Easy / concise wordings, clear meaning.

P8: The buttons are small, sometimes I wasn't sure if I tapped the correct one. Otherwise, I found it quite simple to fill out the survey.

P3: Sometimes it required for me a few seconds more to determine a certain level of a particular emotion, especially the one called being "mentally exhausted", etc. The reason for that was to carefully evaluate influence of other (similar) emotions or states which were listed in the survey. For example "tiredness".

P9: It just required some time, but it was not hard to understand.

Overall Questionnaire: Easiness

Participant	Value	Participant	Value
P01	5	P07	5
P02	5	P08	5
P03	3	P09	4
P04	5	P10	3
P05	5	P11	3
P06	5	P12	5
Total	53		
Average	4,4		
Coding	5		

P6: All surveys (all types) were well designed and easy to fill out. No errors, bugs, etc

P3: The web survey required a bit more time as sometimes I could not recall feeling particular emotion. The iPhone and watch survey solved the problem as I was answering how am I feeling at that moment.

P8: Except for the Apple Watch, where the UI is very small and less responsive, I found it very easy to fill in the surveys.

P10: For the last web diary-questionnaire, I sometimes had my laptop already packed up. Getting the laptop back out again was a little burdensome. However, this is really a minor detail.

Perceived Accuracy

Computer Questionnaire: Influence Question on Answer

Participant	Value	Participant	Value
P01	No	P07	Yes
P02	No	P08	No
P03	Yes	P09	No
P04	No	P10	Yes
P05	No	P11	No
P06	Yes	P12	No
Total	8 No, 4 Yes		
Majority	No		
Coding	No		

P7: Yes because I had to think back how I felt at that time which didn't necessarily correspond to every question asked. For instance when asking about how balanced I was, I had to think about it and then I thought "yes maybe I felt balanced at that time" which wasn't necessarily the feeling I had at the moment.

P10: I was kind of nice. Reflecting on what happened during the day, gave me a good feeling about what I did/achieved that day.

P1: My feelings stayed the same

P4: It is possible. In general, I think the mood was independent form the questionnaire

P6: Again, I think answering these questions in retrospective is always a nice way of thinking how my day has been - and why, also as a trigger to work on oneself.

P3: When I realized I marked quite negative emotions during the whole day, it made me feel slightly miserable.

P8: On the weekend, when I had a guest at my place I felt slightly stressed that I still had to fill in the survey on the same evening. Other than that, I don't think filling in the survey triggered any kind of cognitive process that influenced how I felt (at least not consciously).

P5: It is difficult to specify why, I did not experience a change in my feelings due to answering the questions and thinking about my feelings. Because the feelings I experienced in the evening were usually a mixture of the entire day and/or the current situation and they did not change due to answering these questions. Maybe because I usually filled it out when I was already in my bed and ready for sleeping. ;)

P2: Don’t think so, it did not change the mood during the day

Smartphone Questionnaire: Influence Question on Answer

Participant	Value	Participant	Value
P01	Yes	P07	No
P02	Yes	P08	No
P03	Yes	P09	Yes
P04	No	P10	No
P05	No	P11	No
P06	Yes	P12	No
Total	7 No, 5 Yes		
Majority	No		
Coding	No		

P9: It made me think about my mental state.

P3: Very often, especially regarding the negative feelings, it made me think about the reason why I feel that way.

P7: Since this questionnaire was about how I felt at the time of the survey it was really easy not to be influenced in any way about how I felt. I just straight up entered how I felt.

P5: It did not change my current mood or feelings, because I felt the way I did because of external issues, e.g. deadlines, and a reflection on my feelings did not change anything about them during these days.

P1: For example, when I read the word "stressed" or "in a hurry" I start to feel that.

P8: In the moment it didn't change how I felt. Maybe over time, I became more aware of my emotions because of the conscious process of reporting them regularly.

P6: Again, it makes you think about why you are stressed/anxious/not rested, which is a good thing.

P12: However, I realised my feelings more than otherwise

P2: I guess it helps to be more mindful and aware about the current state. Even though i think the questions are more on the negative side than on the good one.

Smartwatch Questionnaire: Influence Question on Answer

Participant	Value	Participant	Value
P01	No	P07	Yes
P02	No	P08	No
P03	Yes	P09	Yes
P04	Yes	P10	No
P05	No	P11	No
P06	Yes	P12	No
Total	7 No, 5 Yes		
Majority	No		
Coding	No		

P6: Maybe a little bit; It made you stop and think how you actually feel sometimes and why that is. Normally you don't think throughout the day if your anxious e.g., so reflecting about the state of oneself is surely helpful.

P12: It did not influence my feelings, but I realized how I felt

P3: It made me think of particular situations or people which could influence such states/emotions.

P2: I dont think so, but this survey was like a little mindefullness training checking in on the current mood. Which could have further positive effects.

P7: I had a difficult week (emotionally) so I think that the questionnaire probably influenced me because it asked questions that forced me to think about me when I didn't necessarily wanted to.

P9: It made me aware during the day, what my current emotional state was. This allowed me to try and change my outlook if possible.

Computer Questionnaire: Accuracy of response

Participant	Value	Participant	Value
P01	2	P07	4
P02	2	P08	5
P03	4	P09	3
P04	3	P10	4
P05	4	P11	2
P06	3	P12	4
Total	40		
Average	3,3		
Coding	3		

P1: It find it difficult to scale the feelings.

P4: I think my mood of the evening influenced my entries for the different times

P9: Not always sure in the evening how exactly I felt in the morning.

P12: Pretty accurate, sometimes it is hard to distinguish the different kind of feelings because they are similar.

P6: Answering these questions in retrospective is more difficult than ad hoc via app/mobile and I feel it's more a rough estimation. The app/mobile are probably more accurate.

P2: Because of the time difference the answers are not accurate and therefore can’t reflect the feelings as good as the other questionnaires.

Smartphone Questionnaire: Accuracy of response

Participant	Value	Participant	Value
P01	2	P07	4
P02	4	P08	4
P03	4	P09	4
P04	2	P10	4
P05	4	P11	2
P06	5	P12	4
Total	43		
Average	3,6		
Coding	4		

P4: Sometimes I made errors, but could not go back (pushed wrong button)

P8: Sometimes I found it difficult to say whether I am super calm (i.e. a 4) or just calm (i.e. a 3). I tended not to select the extreme answers (1, 4) because I usually felt I could always feel more ... (calm, overwhelmed, stressed, ...).

Smartwatch Questionnaire: Accuracy of response

Participant	Value	Participant	Value
P01	2	P07	4
P02	3	P08	4
P03	4	P09	4
P04	4	P10	4
P05	4	P11	3
P06	5	P12	4
Total	45		
Average	3,8		
Coding	4		

P6: It was a good choice to have a 5-point scale; 3 would be too few (not precise enough), 7 not managable on the hardware.
At least for me, I feel I could very accurately desribe how I feel.

P8: Maybe stronger emotions could be part of the survey too. E.g. "Do you feel depressed / hopeless / energized / excited?".

P12: it is hard to say what the real feeling is

P9: It was nice to be able to answer how I was currently feeling, compared to how I was feeling earlier. This made my responses more accurate i think.

Overall Questionnaire: Accuracy of response

Participant	Value	Participant	Value
P01	3	P07	4
P02	4	P08	5
P03	4	P09	4
P04	3	P10	4
P05	3	P11	2
P06	4	P12	4
Total	44		
Average	3,7		
Coding	4		

P6: As said before: Sometimes I felt that there would be at least one answer missing (big steps between possible answers).

P7: Sometimes it's pretty difficult to really know how we feel (like I am truly stressed, or do I feel balanced right now?)

P4: I think the set of items were quite limited.

P9: I think overall they are quite accurate

P5: This depends on the type of the survey. I felt that the watch and phone were more accurate than the web survey.

P12: I think pretty accurate, because I tried to focus on the questions and distinguish the feelings and answer in a reasonable time

Computer Questionnaire: Way of asking about feelings

Participant	Value	Participant	Value
P01	2	P07	2
P02	3	P08	1
P03	2	P09	1
P04	1	P10	3
P05	2	P11	3
P06	1	P12	1
Total	22		
Average	1,8		
Coding	2		

P3: I do not think it had any impact as I get used to the phrasings of questions and answers.

Smartphone Questionnaire: Way of asking about feelings

Participant	Value	Participant	Value
P01	2	P07	1
P02	4	P08	1
P03	3	P09	1
P04	3	P10	4
P05	1	P11	3
P06	1	P12	1
Total	25		
Average	2,1		
Coding	2		

P3: I had to give a second thought to the questions of being exhausted or tired as I did not want to exaggerate with the first one. Overall, the same answer range kept the survey simple.

P4: as stated above, feeling extremely rested seems strange to me.

Smartwatch Questionnaire: Way of asking about feelings

Participant	Value	Participant	Value
P01	4	P07	1
P02	2	P08	1
P03	4	P09	2
P04	4	P10	3
P05	3	P11	3
P06	1	P12	2
Total	30		
Average	2,5		
Coding	2		

P6: I wouldn't know, why I would be influenced by the wordings. They seemed straight forward to me.

P3: Sometimes all the negative emotions appeared one after another, and then were followed by positive ones. In that cases, it was easier for me to first evaluate the negative states/emotions and then focus on positive aspects (or vice versa). The reason could be connection between some of the emotions, e.g. being tired and anxious, relaxed and rested.

P9: They were simple questions.

P4: Did not like the extremes as they indicated not at all and extremely. I think I ommited extremely s/t more often than others.

OQ: Way of asking about feelings

Participant	Value	Participant	Value
P01	3	P07	2
P02	2	P08	1
P03	1	P09	1
P04	4	P10	3
P05	2	P11	3
P06	1	P12	2
Total	25		
Average	2,1		
Coding	2		

P6: I wouldn't know why.

P10: To be honest, I am not quite sure what this question aims to ask or what I am supposed to answer here.

P4: Again, extremely rested is strange to me. Hence, I did not filled this in.

P9: I would say not at all, I understood all of the questions, so I don't know how they would have influenced me.

P12: Maybe it depended a bit on which feeling was asked first. So maybe I pressed a higher number for „stressed“ and lower for „in a hurry“ because stressed appeared first and in that sense I associated I higher value to that feeling and then to equalize I took a lower value for the second similar feeling

Overall Questionnaire: Missing

P6: I don't think I was unable to respond at any time. Sometimes it just took me longer to figure out how I feel in that given moment. :)

P3: Trying to be as honest as possible, sometimes I realised I marked quite opposite emotions like being balanced and anxious with the same number. It made me think about how I felt and sometimes resulted in slight changes to the survey (only the web one). I believe the reason was trying to hold off the negative emotions, especially during the morning hours, which could influence how I feel during the rest of the day.

P8: I had a busy weekend where I was out and about with friends. When I'm in company I don't check my phone so often or at all, which means I missed some of the iPhone surveys on that weekend.

P7: Once I missed the notification as I woke up too late (shouldn't ask questions over the weekend I think)

P10: Either I was still sleeping, e.g. Sunday morning, or I was unavailable because of a conversation which could either be a meeting or talking to someone on the phone, etc.

P4: had a small number of freezes (no submits); internet connection was not available - not sure whether it submitted the answers later

P9: Often I was not aware there was a new survey available, because it did not vibrate. This was especially the case on the watch, but I'm sure that will be fixed in future studies. If I was in meeting or eating lunch then I would fill out the surveys afterwards

P11: Didn’t notice the notification

P2: Usually i could not repsond when I had to fullfill important tasks like driving a boat or not filling in the survey during a meeting since this is bad meeting ettiquette.

P5: Mainly because I was talking to someone and thought I will answer it later to not disrupt the conversation. This sometimes led to the problem, that I forgot to answer it at all. A typical situation was a meeting with someone or also just a small coffee break.
Another reason was, that I got the notification while I was on my way or catching a bus and therefore could not answer it.

P12: Sometimes I did not respond immediately, so that was because of having meetings at work or not seeing that a new survey was opened

P1: The phone questionnaire didn't work in the first day and didn't get the notifications. I got some notifications during my meetings and was not able to fill out the questionnaires.

Overall Questionnaire: Representative Reason

Participant	Value	Participant	Value
P01	2	P07	5
P02	5	P08	5
P03	4	P09	3
P04	4	P10	5
P05	4	P11	2
P06	4	P12	5
Total	48		
Average	4,0		
Coding	4		

P6: Generally, yes. Maybe I was a little bit more often alone in the office as my co-worker was partly gone for vacation. :)

P3: Not necessarily, as I gave up the evening resting time in order to work and study for the exams. I believe this could affect my emotions.

P8: Yes, these were pretty typical days for me.

P4: Every day is quite different. But I think there might be some systematic weekend effects :)

P9: I'm slowly going towards exam time, so I am more stressed then usual but its not full blown exam panic time yet.

P2: of my work live yes I have the feeling i was always at work when i had to fill in the survey.

P12: Normal working days and private appointments as usual

Effect Device Type

Computer Questionnaire: Effect Device Type

Participant	Value	Participant	Value
P01	3	P07	1
P02	3	P08	2
P03	2	P09	4
P04	2	P10	3
P05	1	P11	3
P06	1	P12	2
Total	27		
Average	2,3		
Coding	2		

P1: My answers are less spontaneous doing the questionnaire on computer.

P4: Cannot imagine how - perhaps that I could change my answer was a good thing

P9: Yeah it seemed to take longer than on the phone because you need to fill out the same stuff for each time period, instead of actually filling it out directly during the day. Made it pretty repetitive.

P12: Maybe a bit, because several questions appear on one page, so, one can scroll pack and rethink the answers and compare the answers for the different feelings.

P6: I would not know why.

P3: I believe seeing the choices I made, e.g. being tired, anxious could slightly influence my further choices.

P8: See the answer to "On a scale from 1 to 5, do you think the questions were difficult to understand?" → I made fewer "wrong-side-of-the-scale errors", i.e. thinking that 1 was the "positive" answer when in fact, depending on whether the question asked about a positive emotion (calm, balanced...) or a negative one (overwhelmed, stressed, anxious...), 5 was the "positive" side of the scale.

P2: difficult to say... It is much easier at the computer to type in long texts which would be very burdensome on the watch or on the phone, but since most parts of the survey was just clicking on numbers this did not change much. (But I did one final survey on the phone which was very annoying and which led also to much shorter answers. Because of the annoyance of typing everything on the phone which also can make me angry ;))

Smartphone Questionnaire: Effect Device Type

Participant	Value	Participant	Value
P01	2	P07	1
P02	3	P08	1
P03	1	P09	2
P04	2	P10	1
P05	4	P11	2
P06	1	P12	1
Total	21		
Average	1,8		
Coding	2		

P9: I think because it was so easy to answer, i was less annoyed.

P5: In the last feedback survey (the one about the watch), I answered that I missed a neutral position. Now I noticed, that I might had insecurities because of the word 'moderately'. I interpreted it as 'ziemlich' and therefore, the 'distance' between 'not at all' and 'moderately' was for me quite large.

P6: I wouldn't know why the modality would be influential.

P2: Not clear problematic was that there was no back button, so sometimes wrong awnsers could not be corrected.

Smartwatch Questionnaire: Effect Device Type

Participant	Value	Participant	Value
P01	2	P07	1
P02	2	P08	1
P03	1	P09	5
P04	3	P10	3
P05	4	P11	1
P06	1	P12	1
Total	25		
Average	2,1		
Coding	2		

P5: There is no neutral option, so I always have to answer a tendency towards the positive or negative. I was wondering, whether a slider as an answer option would be possible and if it would change my answers.

P2: Maybe a little bit but this might be mostly because of the scale of the answers. A 4 scale answer is easier to decide on than a 7 level answer.

P9: I think I was quicker to answer, if the watch vibrated (which it din't always do).

P4: AS scales were shorter I can imagine some tendencies to the extremes.

Overall Questionnaire: Comment

P6: Thank you for your invitation to the survey / testing the app! I liked it a lot. And good luck with your research

P10: It feels like there are some repeating questions.
Maybe I am wrong and it just feels like it. However, it would be nice to be able to go back to check if the questions were actually similar.
As there are a lot of free text entry possibilities, I say something in one field that could also go into a different one.

Reactivity to Notification

Participants responded to their notification on their iPhone after 30 minutes on average. The average reactivity to the notifications decreased from 40 minutes on the first day to 25 minutes on the last day. The fastest was one participant which reacted to the notifications over a whole day on average in 14 seconds. It took participants up to 3 hours to respond to the notification on their iPhone.

Participants responded to the notifications on their Apple Watch in an average of 20 minutes. The average reactivity over all participants decreased from 21 minutes on the first day to 12 minutes on the last day. The analysis of the reactivity can be found on the CD-ROM in the folder AboutyourDay.zip.

Appendix E: Content of the Enclosed CD-ROM

Abstract.txt	Abstract of the thesis
Zusfsg.txt	German abstract of the thesis
Master Thesis.pdf	This document
AboutyourDayApp.zip	Zip-File containing the application source code.
AboutyourDay.zip	Zip-File containing the raw data and R code for analysis.