MOBILE USER DATA MINING TO INFER KNOWLEDGE WORKERS' DIFFERENCES IN OFFICE ENVIRONMENTS FOR EFFECTIVE HEALTH INTERVENTION DELIVERY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

ŞEYMA ÇAVDAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

AUGUST 2019

MOBILE USER DATA MINING TO INFER KNOWLEDGE WORKERS' DIFFERENCES IN OFFICE ENVIRONMENTS FOR EFFECTIVE HEALTH INTERVENTION DELIVERY

submitted by **ŞEYMA ÇAVDAR** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics** ________________

Prof. Dr. Yasemin Yardımcı Çetin
Head of Department, **Information Systems** ________________

Assoc. Prof. Dr. Tuğba Taşkaya Temizel
Supervisor, **Information Systems, METU** ________________

Assoc. Prof. Dr. Pekin Erhan Eren
Co-supervisor, **Information Systems, METU** ________________

Examining Committee Members:

Prof. Dr. Özlem Bozo
Psychology, METU ________________

Assoc. Prof. Dr. Tuğba Taşkaya Temizel
Information Systems, METU ________________

Assoc. Prof. Dr. Banu Günel Kılıç
Information Systems, METU ________________

Assist. Prof. Dr. Bilgin Avenoğlu
Computer Engineering, TED University ________________

Assist. Prof. Dr. Gönenç Ercan
Computer Engineering, Hacettepe University ________________

Date: 28.08.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Şeyma Çavdar

Signature :

ABSTRACT

MOBILE USER DATA MINING TO INFER KNOWLEDGE WORKERS' DIFFERENCES IN OFFICE ENVIRONMENTS FOR EFFECTIVE HEALTH INTERVENTION DELIVERY

Çavdar, Şeyma

Ph.D., Department of Information Systems

Supervisor: Assoc. Prof. Dr. Tuğba Taşkaya Temizel

Co-Supervisor: Assoc. Prof. Dr. Pekin Erhan Eren

August 2019, 148 pages

Owing to the widespread and ubiquitous nature of mobile technologies, a large amount of data about users including location, access and interaction behavior is currently available. This data has recently become important as it has the potential to reveal personal information, social context and user characteristics, which can be significant for effective health interventions through mobile phones. Accordingly, this thesis mainly aims to explore the individual differences of knowledge workers and social context in order to infer their available moments using mobile sensor data. A hybrid personalized model is presented as a novel approach for this purpose. Based on the model results, it is found that time, location characteristics, ringer mode, and user activity are effective in predicting availability. In addition, it is investigated how knowledge workers' engagement/challenge levels during work hours are related to their personality traits, social norms in office environments, and mobile application usage. The results show that personality traits and mobile application usage during work hours are significantly related to the engagement and challenge levels, however, social norms have a marginal effect on them. The results of the study present valuable implications for further studies and mobile application designs, which aim to understand the individual differences of employees in office environments.

Keywords: Mobile Health Interventions, Mobile User Modelling, Data Mining, Office Norms, Rest Breaks

ÖZ

ETKİLİ SAĞLIK MÜDAHALELERİ GÖNDERİMİNDE OFİS ORTAMINDAKİ ÇALIŞANLARIN FARKLILIKLARINI ANLAMAK İÇİN MOBİL KULLANICI VERİSİ MADENCİLİĞİ

Çavdar, Şeyma
Doktora, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Doç. Dr. Tuğba Taşkaya Temizel
Ortak Tez Yöneticisi: Doç. Dr. Pekin Erhan Eren

Ağustos 2019, 148 sayfa

Mobil teknolojilerin yaygın ve her yerde kullanımı sebebiyle, günümüzde kullanıcılar hakkında yer, erişim ve etkileşim davranışları gibi fazla sayıda veriye erişmek mümkündür. Bu veriler, mobil telefonlar aracılığıyla etkili sağlık müdahaleleri gönderiminde anlamlı olabilecek kişisel bilgileri ve kullanıcı karakteristiklerini ortaya çıkardığından son zamanlarda önemli duruma gelmiştir. Bu tez çalışması temel olarak mobil sensör verisi kullanarak ofis çalışanlarının uygun vakitlerini anlamak üzere kişisel farklılıklarını ve sosyal bağlamı araştırmayı hedeflemektedir. Bu amaçla, yeni bir yaklaşım olarak hibrit kişiselleştirilmiş bir model sunulmaktadır. Model sonuçlarına göre, uygun vakitlerin tahmininde zaman, mekan özellikleri, telefonun ses modu ve kullanıcı aktivitesi önemli bulunmuştur. Ayrıca ofis çalışanlarının iş meşguliyet ve zorluk seviyelerinin kişilik özellikleri, ofis ortamındaki sosyal normlar ve mobil uygulama kullanımı ile ilişkisi araştırılmıştır. Sonuç olarak, iş meşguliyet ve zorluk seviyesinin kişilik özellikleri ve mobil uygulama kullanımı ile önemli derecede ilgili olduğu; sosyal normlarla olan ilişkinin ise daha az önemli olduğu bulunmuştur. Çalışmanın sonuçları, ofis çalışanlarının çalışma ortamlarındaki farklılıklarını anlamayı amaçlayan ileriki çalışmalar ve mobil uygulama tasarımları için değerli çıkarımlar sunmaktadır.

Anahtar Kelimeler: Mobil Sağlık Müdahaleleri, Mobil Kullanıcı Modellemesi, Veri Madenciliği, Ofis Normları, Dinlenme Araları

To my beloved husband, Davut,
To my little son, Umut,
To my mother, Muazzez,
To the memories of my father, Osman, & of my grandmother, Hidayet…

ACKNOWLEDGMENTS

First of all, I would like to express my sincere appreciation to my supervisor, Assoc. Prof. Dr. Tuğba Taşkaya Temizel, for her endless support, motivation, valuable guidance, and encouragement. It was a great pleasure and honor to have her supervise me. I appreciate all her contributions during my Ph.D. study to make this work fruitful and stimulating.

I am thankful to Prof. Dr. Mirco Musolesi, Prof. Dr. Peter Tino, and Dr. Abhinav Mehrotra for their valuable advice and contributions to my study. I am also grateful to my co-advisor Assoc. Prof. Dr. Erhan Eren and to my thesis monitoring committee members, Prof. Dr. Özlem Bozo and Assoc. Prof. Dr. Banu Günel Kılıç, for their valuable feedbacks. I have learnt a lot from all of them. I would like to thank the examining committee members, Assist. Prof. Dr. Bilgin Avenoğlu and Assist. Prof. Dr. Gönenç Ercan.

I would like to thank the participants of the pilot study, particularly, Nurcan Alkış, Cihan Öngün, Didem Tokmak, Samet Nargül, and Üzeyir Doğan, for their patience for providing additional data in every experiment day. I also thank Erkam Akkurt, Özlem Türkeri, and Mehrdad Mizani for their help in designing the experiment. I am also deeply thankful to the participants of the study, this study could not be completed without their participation.

I acknowledge the support of The Scientific and Technological Research Council of Turkey (TÜBİTAK) BİDEB for 2211-E graduate student fellowship during my Ph.D. education.

I would like to thank all of my colleagues and friends, Özge Gürbüz, Serhat Peker, Ece Işık Polat, Nur Didem Başkurt, Ali Mert Ertuğrul, Gökçen Yılmaz, Duygu Fındık Coşkunçay, and Sibel Gülnar. I shared many great moments with them and they made my Ph.D. study enjoyable.

I am deeply grateful to my mother, Muazzez Küçüközer, for her unconditional love. She always believed in me and she was there all the time whenever I needed her. I would not have completed this study without her endless support.

I am thankful to my son, Umut Berk Çavdar. He is the most valuable gift that this life has given to me. He made me feel alive with his smiles and hugs, and they were priceless; it is never enough how much I thank him.

Last but surely not the least, I would like to express my sincere gratitude to my beloved husband, Davut Çavdar, for his never-ending love, valuable support, and continuous patience. He always held his head high regardless of what and his courage gave me the strength I needed not only during my Ph.D. study but also during the lifetime we shared. I am so glad to have him in my life. I believe that we can overcome everything in this life together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
AP	Access Point
API	Application Programming Interface
AU	Application Usage
BCSS	Behavior Change Support System
BHI	Behavioral Health Interventions
BPTI	Basic Personality Traits Inventory
BSSID	Basic Service Set Identifier
dBCIs	Digital Behavior Change Interventions
DIC	Deviance Information Criterion
ESM	Experience Sampling Method
GCM	Google Cloud Messaging Service
GLMM	Generalized Linear Mixed Model
GPS	Global Positioning System
k-NN	k-Nearest Neighbor
KDE	Kernel Density Estimation
LF	Location Frequency
LS	Location Similarity
MCMC	Markov chain Monte Carlo
METU	Middle East Techical University
MPPUS	Mobile Phone Problem Use Scale
MRS	Middle Response Style

NB	Naive Bayes
NOA	Number of Applications
NOS	Number of Switches
NRS	Negative Extreme Response Style
PRS	Positive Extreme Response Style
PSD	Persuasive Systems Design
RM	Ringer Mode
RMC	Ringer Mode Change
rmcorr	Repeated-measures correlation
RSSI	Received Signal Strength Indicator
SSID	Service Set Identifier
SVM	Support Vector Machines
TSL	Time Spent in Location
Wi-Fi	Wireless Fidelity
WRMSDs	Work Related Musculoskeletal Disorders

CHAPTER 1

INTRODUCTION

Most of the daily activities can be performed via mobile devices such as smartphones and tablets. People carry their mobile devices with themselves almost every moment in their lives and their first communication channel with other people have become their mobile devices. Various tasks such as sending e-mail, checking social media, chatting, and following up events on digital calendars could be handled via those devices. One of the prominent areas that mobile devices are currently being widely used is health domain. Most of the health-related information can be kept on mobile health applications such as daily steps taken, heart rate, keeping diaries about meals, or recording medication intake. Smartphones enable people to make use of almost all technological opportunities. Since such availability is present on mobile devices, several applications use reminder feature of smartphones in order for people to perform health-related behaviors such as reminding to take medication, to take a walk, or to drink water. Those are simple examples of health interventions, which can be described as "interventions designed to affect the actions that individuals take with regard to their health" [3].

Mobile health systems make use of persuasive technologies when delivering interventions to users in order to be more effective. Persuasive technologies are defined as "interactive information technology, which aims to alter users' behaviors or attitudes" [4]. As Fogg stated in his Fogg Behavior Model [5], appropriately timed triggers should be present in order to alter or change the behavior. That brings us to finding an appropriate time for mobile notifications that deliver health intervention related messages to users. How could those appropriate moments be identified? In fact, this question has been widely investigated by a dozen of studies. As a starting point, Oinas-Kukkonen [6] state that context information is necessary to determine "opportune moments" before delivering interventions.

Several works have been made on inferring opportune moments via mobile phones, wearable sensors, desktop computer use etc. Those studies focus on not only sending health interventions but also sending intelligent mobile notifications in order for them to be accepted by users and not to interrupt them. Based on the general findings of those studies, interruptibility, which refers to the "ongoing status of a user with regard to receptivity to get messages" [7], are affected by several factors which include time, location of the user, physical activity, ringer mode of the mobile device, application usage, social engagement, and level of focus.

One of the important health risks in today's world is sedentary life style. Several studies show how a sedentary life style causes metabolic problems [8, 9]. One of the populations that has a sedentary life style is office workers since they accomplish their tasks using computers and sitting on their desk. Many office workers in today's world face repetitive movements and/or static postures in their work lives. That causes several work-related musculoskeletal disorders (WRMSDs) such as carpal tunnel syndrome and repetitive strain injury. Prevention from WRMSDs includes simple micro rest breaks, which last for 30-60 seconds at every 15-20 minutes [10–15]. It is also suggested to perform simple stretching exercises or to walk [16, 17]. However, most office workers may forget to take even the micro breaks from their repetitive works due to focusing on their work, the lack of motivation, or the unawareness regarding the importance of those breaks.

Behavioral health interventions mentioned above could be very effectively used for reminding office workers to take rest breaks. Similarly, the timing problem occurs for delivering interventions because an office worker may not take a break in the middle of a meeting, or when s/he has a high level of focus on the ongoing task. Therefore, it is important to develop a solution, which may be used when users are more inclined to have rest breaks and accept notifications without an interruption. In a few studies, office workers' interruptibility has been explored (e.g. [18]). However, those studies focus on general interruptibility whereas availability for taking a rest break differs from interruptibility in terms of the duration. In some cases of rest breaks, longer time is needed to obtain one's attention rather than a quick response. In such situations, an office worker might be interruptible for that specific moment however s/he might not be interruptible for the following 5 or 10 minutes for performing exercises or taking a short walk.

Another problem in the current literature, is related to building individual models. Individual models, as can be understood from the term itself, are the models for the inference of a target variable (e.g. interruptibility) developed upon a single user's data. Individual models enable to infer each participant-specific situation since interruptibility or available moments are quite personalized terms. On the other hand, generalized models, which are built on the whole users' data, give population-specific results, and may not be applicable for all users in fact. Previous studies show the efficiency of individual models (e.g. [18, 19]) when the data points for an individual model exceed approximately 50 [18]. When there is a number of data points less than 50, individual models may fail to learn the characteristics of the user, which is named as "cold-start" problem [18].

When the two issues mentioned above (the need for an inference of taking rest breaks in a specified duration and the cold-start problem) are taken together, there is a need for developing a solution for *inferring the available moments of office workers for taking rest breaks with the consideration of individual differences among the workers*.

When it comes to office workers, the main factors that should be considered in the designs include their work engagement and challenge levels, as well as their attentional states. Those metrics could be assessed with context information. For example, an increase in smartphone usage may be a sign of boredom since most users prefer using mobile phones when they feel bored [20–22]. Similarly, their interaction with computers (which programs or how long they use) also reveals boredom [2, 23]. The moments when they feel bored from their work, i.e. they do not feel challenged or engaged with their work, could be quite appropriate moments to deliver health-related notifications. However, when modelling individual engagement/challenge levels or attentional states, the cold-start problem also may occur.

How could *in-situ* engagement/challenge levels be obtained from the office workers, and how would the responses of office workers vary? This brings us to the topic related to responsiveness. In fact, responsiveness is commonly used as the synonym of receptivity or receptiveness even in the interruptibility studies, which capture in-situ behaviors generally with Experience Sampling Method (ESM). However, responsiveness in the relevant literature has always been tackled in one dimension: whether a user responded to a survey, notification etc. or not. Although users may appear to be attentive to the surveys at first, their answers might be inaccurate, repetitive or random. Hence, response style of users should be understood in order to make more reliable inferences. Response style is defined as "a respondent's tendency to responding systematically to questionnaire items regardless of the content" [24].

Responsiveness to health-related notifications can be affected by the health history of users. It is well known that discomfort caused by a disorder determines the level of adherence to treatment [25]. In the mobile context, users may be more likely to respond to messages sent by break reminder applications, if they experienced musculoskeletal discomfort due to their sedentary life. Awareness about the consequences of their actions is also important. A recent study showed the effects of self-regulation and habit strength on the sedentary behaviors of knowledge workers [26]. Higher awareness or self-regulation may be an indicator for the responsiveness to break-reminder notifications. Besides, social factors such as subjective norm have been found as a precursor related to behavioral intention [27,28]. Subjective norm is simply a perception towards performing a behavior influenced by others who are important to the one performing the behavior [29]. Recent studies showed that office employees are influenced by their co-workers regarding prolonged sitting behavior [30] or performing physical activity [31]. Hence, office workers might also be influenced to take rest breaks by their colleagues. Finally, the number of colleagues in the same office might be another factor for both responsiveness and work engagement/challenge levels. It has been shown that office type (shared or private offices) has a significant effect on distractions [32] and also on sitting time [33].

Due to the reasons stated above, there is a need for *understanding which personal or social factors affect the responsiveness of office workers in terms of response style metrics*. Besides, considering individual differences, a model which is able to deal with a low number of data points, should be developed for the inference of engagement/challenge levels and attentional states of office workers since those are quite useful information for the inference of available moments. Therefore, based on the current gaps in the literature, the main objectives of this thesis can be described as below:

- To build a novel method for inferring the rest break availability of office workers with the consideration of cold-start problem and repeated-measures design of the data set,
- To investigate which mobile sensor data is important for the inference of rest break availability in office settings,
- To investigate the relation between the responsiveness and break-reminder notifications, personality, office-related factors, mobile application usage, awareness and musculoskeletal discomfort of office workers,
- To build a model for inferring in-situ attentional states and engagement/challenge levels of office workers with the consideration of cold-start problem, the variety in the number and the characteristics of the responses, and repeated-measures nature of the data,
- To investigate which application usage metrics are important for the inference of in-situ attentional states and engagement/challenge levels.

1.1 Research Questions

The following research questions are developed with respect to the objectives of the study:

1. How can a model be built for inferring availability of office workers for having rest breaks using mobile phone sensors by considering cold start problem, the variety in the number and characteristics of the responses, and repeated-measures design of the data? How is this model comparable to individual and general models?

2. How are the musculoskeletal discomfort, awareness, office-related factors, personality traits, and mobile application usage of office workers related to their responsiveness to break-reminder notifications?

3. Which application usage metrics are related to in-situ engagement/challenge levels of office workers?

4. How can a model be built for inferring attentional states and engagement/challenge levels of office workers using application usage metrics by considering cold start problem, the variety in the number and characteristics of the responses, and repeated-measures design of the data? How is this model comparable to individual and general models?

In this thesis, a hybrid model is proposed for analyzing the availability of office workers for taking rest breaks during work hours with mobile phone sensors as a solution to the first research question. In order to analyze the second research question, a research framework is proposed for understanding the responsiveness of office workers to the engagement/challenge questions. The responsiveness metrics are hypothesized to be affected by awareness of rest breaks, musculoskeletal discomfort, personality traits, mobile application usage, and office-related factors. For the third and the fourth research questions, the application usage effects on *in-situ engagement and challenge levels* are explored with the consideration of repeated-measures design.

In order to validate the models and the framework proposed in the thesis, a user experiment was designed. The experiment was conducted with 31 participants in 10 workdays during their work hours. Firstly, a survey was conducted among office workers. The factors related to musculoskeletal discomfort, awareness, and office-related factors were collected with this questionnaire. Secondly, the participants installed the mobile application, which was developed for delivering break reminder notifications and collecting context data. Several break-reminder notifications were sent to the participants via the mobile application. *In-situ* engagement/challenge levels and break availability of the participants were collected with the experience sampling method (ESM) questions, which were sent with the reminders. The application also provided exercises, which can be taken during users' rest breaks. Then, the participants filled personality traits inventory and problematic mobile phone usage inventory during the experiment. Finally, after the experiment ended, the participants filled the post-experiment questionnaire, which assesses the usability of the application.

In total, 528 valid ESM responses were collected from the participants, whose answer rate was above 25%. Those responses were used for validating the hybrid model presented as the first study. The hybrid model works in two-phases: First, the kernel density estimates of each user's self-reported break availability response with its corresponding timestamp were calculated. Then, the availability of office workers was modelled with the features selected (kernel density estimates, location parameters, ringer mode parameters, physical activity, and mobile application usage). In this phase, Generalized Linear Mixed Model (GLMM) using a Markov chain Monte Carlo method was employed since the data set of the study had a repeated-measures design. GLMMs incorporate fixed and random effects together, so it enables to fit model parameters on both population- and individual-level. In similar studies, classification methods such as random forest models or support vector machines are generally used for modelling all participants' data, and they were named as general models [18,19,34]. However, with such approaches, the relation among measurements from the same participant is ignored, and the assumption of statistical independence is violated. As a remedy, individual models are built for each participant. Yet, this time, a significant amount of data is required for each individual model in order to make them work effectively. In this study, the validation of the analyses using different sub-samples of the data set and a comparison of GLMM with general and individual random forest classifiers are presented.

In the second study, the responsiveness is investigated with several metrics such as acquiescence, dis-acquiescence, and extreme response style (negative and positive). The factors of personality traits (Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness, and Negative Valence), mobile application usage, office-related factors, musculoskeletal discomfort, and awareness about having rest breaks are hypothesized to be effective on the responsiveness. Correlation analyses are performed among the variables, and significant relations are presented as a result.

In the third study, the relation between in-situ engagement/challenge levels and mobile application usage is explored. Context information such as time and location of individuals and the activity information have an impact on inference on engagement and challenge levels as shown in [35, 36]. However, this study is focused solely on inference using mobile application information and statistics derived from this data. Unlike the previous studies, a recently proposed correlation metric called repeated-measures correlation, which is designed specifically for repeated-measures studies [37] is performed. Both the short-term application usages (e.g. 5-10-15 minutes) and the long-term application usage (e.g. 30-45-60 minutes) are analyzed for the inference of engagement/challenge levels. Finally, engagement/challenge levels and attentional states of office workers are modelled with GLMMs.

1.2 Contributions of the Study

The main contributions of the thesis are given as follows:

- A novel method is proposed for the inference of rest break availability of office workers as a remedy of analyzing unbalanced (in terms of users' responses) and limited data.
- Generalized linear mixed model using a Markov chain Monte Carlo method is used for modelling break availability, attentional states, and in-situ engagement/challenge levels. The main advantage of GLMM resides in its ability to address both within and between subjects' factors successfully. Hence, it might be used as a solution to the cold-start problem of individual models when there is unbalanced limited data to predict a user's availability.
- Repeated-measures correlation analysis is used as different from the previous studies, in which mainly simple correlation techniques on the aggregated data are used, for investigating the relation between mobile application usage and in-situ engagement/challenge levels.

- New features are proposed to represent the relation between individuals and the indoor locations they are in for modelling user's availability for taking breaks.
- A new research framework is proposed for the inference of engagement/challenge levels of office workers and their responsiveness to health-related mobile notifications.
- Responsiveness is investigated with several metrics such as acquiescence, disacquiescence, extreme response style (negative and positive), which have not been used in the interruptibility or work engagement/challenge studies.
- A population-specific personality inventory is used in the study, which has not been used before in office settings.

1.3 Organization of the Thesis

The thesis includes six chapters. Chapter 2 presents the literature review including interruptibility, mobile interventions in health domain, persuasive and anticipatory systems, work engagement/challenge and attentional states of office workers, work-related musculoskeletal disorders and sedentary behavior in office environment, subjective norm in office setting, and responsiveness and variability measures.

Chapter 3 presents the user experiment in detail. Research method and experiment design, mobile sensing application and instruments used in the study, reminder messages, pilot study, participants information, and data collection procedure are given in this chapter.

Chapter 4 introduces a hybrid model for predicting office workers' rest break availability using mobile sensors. The model is two-staged: in the first phase, time information is used, then, in the second phase, the time information processed in the first phase and other context information are used for predictions both at individual-level and population-level.

Chapter 5 presents the analyses regarding the responsiveness about work engagement and challenge questions. In this chapter, the social and personal indicators for predicting the responsiveness of office workers are analyzed, and the results are presented.

In Chapter 6, the relation between mobile application usage parameters and in-situ attentional states, engagement/challenge levels of office workers with the repeated-measures analysis techniques.

Finally, Chapter 7 concludes the thesis with the contributions and practical implications, and the future work is presented.

CHAPTER 2

RELATED WORK

In this chapter, related studies are given in detail. In the first section, the studies related to opportune moments and interruptibility context are explained. Then, the details of behavioral health interventions are presented. The studies related to work engagement, challenge and attentional states in work places are given in the third section. Then, the studies in regard to understanding office context are presented. The chapter ends with the implications from the studies discussed.

2.1 Opportune Moments and Interruptibility Context

In recent years, with the emergence of mobile and ubiquitous computing, there has been an increase in the field of mobile interruptions in order to use mobile devices (e.g. smartphones) more effectively. An interruption can be defined as an "external random event that diverts a user's attention away from the current task cognitively" [38, 39]. Several researchers conducted studies to understand the interruptibility of users, which refers to "the ongoing status of a user with regard to receptivity to get messages" [7]. Preliminary works have been conducted in desktop settings. As mobile technologies have increased, the focus of the researchers has switched to this area.

Opportune moments or timing of interruptions are important because users are most likely to ignore a message (or notification) when they are busy even if the message is important. In order to reach or attract user, the message should be delivered at the right moment. Especially, in the behavior change or persuasion domain, the timing is essential to deliver persuasive messages to users.

As Oinas-Kukkonen [6] state in his Persuasive Systems Design (PSD) model, context information is necessary to determine opportune moments before delivering interventions. So, in order to understand which moments are appropriate to deliver a message to users, users' context should be identified. There may be several factors in user's context that affect the perception of interruptions. Ho and Intille [1] described these factors, and many of the subsequent studies guided these 11 factors described. These factors and their definitions are given in Table 1.

Grandhi and Jones [40] presented a framework for interruptibility. They also emphasized the impact of "relational context" in order to determine interruptibility or manage interruption. They separated the interruption context into three categories:

- *Cognitive context:* All elements that cover user's cognitive engagement in tasks and the effects of interruptions on task performance.

Table 1: User context factors affecting the perception of interruptions [1]

Factor	**Explanation**
Activity of the user	Activity that the user is busy with during the interruption
Utility of message	Perceived importance of the interruption by the user
Emotional state of the user	Mentality of the user, and the relation the user has with the interruption medium
Modality of interruption	Medium of the interruption delivered
Frequency of interruption	Rate of interruptions occurring
Task efficiency rate	Time that takes to understand the interruption task and the duration of the task to be completed
Authority level	Control of the user over the interruption medium
Previous and future activities	Activities that the user was occupied with before the interruption and the ones which might be occupied in the future
Social engagement of the user	Role of the user in the ongoing activity
Social expectation of group behavior	Reaction of nearby people to the interaction
History and likelihood of response	Pattern that the user performs with regard to the interruption

- *Social context:* All elements that cover user's perceived physical environment in terms of sociality such as where interruptions come from, other individuals in the environment and their relationship with the user, the social activity in the environment.
- *Relational context:* All elements that cover user and the interruption such as the relationship between the user and the interrupter, the content of interruption, the conditions of interruption, the history of interaction between the user and the interrupter.

There are several studies that show the importance of context information on the interruption management. In Appendix A.1, current studies about interruptions and opportune moments are summarized. The context information that was found significant in predicting opportune moments or on the efficacy of interruptions are given.

As a result, it has been observed the following conclusions from the current studies in the literature:

- *Timing* of an interruption is an important factor on the receptivity of users. Users more likely to respond to the messages sent at opportune moments than the ones sent at random times.
- Besides the timing of the interruption; *content* of the interruption, *location* of the user, *application type* that produces the interruption, *perceived importance* of the interruption, *activity* of the user, *social engagement* of the user, *ringer mode* of the mobile phone and *application usage* prior to notifications, *level of focus* are other factors that affect the responsiveness of users to interruptions.
- In recent years, interruptibility studies focus on building *personalized* models, but they can suffer from the *cold-start problem*, which is the lack of sufficient individual data for training a model in the beginning.

2.2 Behavioral Health Interventions (BHI)

In this section, the theories related to behavioral health interventions are described first. Then, digital interventions in health domain are presented. Specifically, persuasive systems, anticipatory mobile computing, mobile interventions for behavior change, and messages used in those interventions are discussed in detail. The section ends with the details of a more specific domain for behavior interventions as work-related musculoskeletal disorders and sedentary work styles of office workers.

2.2.1 Theoretical Background for BHI

Behavioral intervention can be described as "interventions designed to affect the actions that individuals take with regard to their health" [3]. There are many factors that affect health behaviors of individuals such as social, cultural, or economic factors. Attitudes, reactions, motivation, and knowledge are among the most important antecedents of health behaviors, then, social factors such as social relationships or culture are second important determinants [41]. Hence, it is a complex task to define the determinants of a health behavior. However, it is acceptable to think from a social point of view, instead of just from the individual side [41].

There are several theories related to understand and define the health behaviors of individuals. The most prominent theories used in the studies recently, can be described as Health Belief Model, Transtheoretical Model (Stages of Change), Social Cognitive Theory, and Social Ecological Model [41]. Below, a brief summaries of those theories are presented:

- **Health Belief Model:** It was developed in 1950s to understand and predict health-related behaviors, specifically regarding why people use or not use public health services [42]. Later, the model has evolved to cover other health-related issues such as injury prevention [43]. The key concepts of the Health Belief Model are perceived susceptibility, perceived severity, perceived benefits, perceived barriers, cues to action, and self-efficacy [43].

- **Transtheoretical Model / Stages of Change:** Prochaska and DiClemente [44] introduced the Transtheoretical Model which suggests that people move among five stages when changing or modifying their behavior. The five stages of the model are pre-contemplation, contemplation, preparation, action, and maintenance. Several health-related behaviors such as smoking, physical activity, and eating habits could be explained with the Transtheoretical Model (e.g. [45]).

- **Social Cognitive Theory:** It comes from the social learning theory of Bandura [46], and human behavior is explained by personal and environmental factors in Social Cognitive Theory [47]. Social Cognitive Theory claims that people learn via observing other people's behavior and the results of those behavior [46, 47]. The main concepts in the theory are observational learning, reinforcement, self-control, and self-efficacy.

- **Social Ecological Model:** In this model, the levels of influence are emphasized as individual, interpersonal, organizational, community, or public policy [48]. Based on the model, behaviors both affect and are affected by the social environment. The key principles of the model are in line with the concepts of Social Cognitive Theory. Basically, the model emphasizes the importance of the environment on health-related behaviors.

2.2.2 Digital Interventions in Health Domain

In this section, persuasive systems (behavior change support systems) are explained first. Secondly, the importance of personalization in persuasive systems is presented. In the third subsection, anticipatory mobile computing regarding behavior change and the responsiveness in those systems are given. In the fourth subsection, the sample studies related to mobile interventions for behavior change are presented. Finally, the message types sent in those interventions are given.

2.2.2.1 Persuasive Systems (Behavior Change Support Systems)

Persuasion is defined as changing behaviors or attitudes of others toward a system, an idea or a person. Persuasive technology has been known as interactive information technology, which aims to alter users' behaviors or attitudes [4]. Accordingly, persuasive systems are defined as "computerized software information systems designed to reinforce, change or shape attitudes or behaviors or both without using coercion or deception" [49]. The terms of "persuasive systems" and "behavior change support systems (BCSS)" are used synonymously in the literature.

Fogg [4] is the first one who stated that information technologies could be used for persuasion. He developed a framework named "Fogg Behavior Model", which consists of three primary elements: motivation, ability and triggers [5]. According to this framework, in order for a person to perform a target behavior, one needs to have a high motivation and a high ability. In addition, appropriately timed triggers should be present because a person may have high ability and high motivation but something is needed to trigger that behavior to occur. In this framework, the timing of triggers is important as described in Section 2.1.

Oinas-Kukkonen and Harjumaa [49] summarized the key approaches previously defined for human-computer persuasion. These are information processing theory [50], cognitive consistency theory [51], elaboration likelihood model [52], influence techniques approach [53], coactive approach to persuasion [54], and persuasive technology framework [55]. Then, they have introduced a framework named "Persuasive Systems Design", which consists of three major steps: Understanding key issues of persuasive systems, analyzing persuasion context, and design of system qualities [56]. These steps lead to behavior and/or attitude change. Timing of message delivery is also highlighted by Oinas-Kukkonen [6]: According to their PSD model, persuasion context should be carefully analyzed in order to recognize opportune moments.

The concept of BCSS was first introduced by Oinas-Kukkonen [57], and is based on the PSD model. BCSS are widely used in the health domain for users to attain their health-related goals such as smoking cessation, alcohol abuse, obesity, diabetes, asthma, stress, anxiety, depression, and insomnia [58]. Oduor et al. [59] proposed software design patterns for BCSS. The four patterns presented are social learning and facilitation, competition, cooperation, and recognition. Social learning means that users observe others' behaviors through persuasive systems, and social facilitation means that users realize that other peers also perform the target behavior with them. Competition which is also presented in the PSD model, motivates users to adopt target behavior with the feeling of competitiveness, whereas cooperation motivates users with leveraging reciprocity behavior of the human nature. Finally, recognition pattern stands for the system capability that increases the likelihood of adopting a target behavior by proposing recognition by other users or social groups.

2.2.2.2 Personalization in Persuasive Systems

Personalization in persuasive systems is important and makes BCSS more effective [60–63]. The common statement in the studies is that individuals differ from each other, and each has different level of persuadability. Hence, persuasive systems should adapt themselves to different types of individuals, and persuasive approach for each individual should be different.

Kaptein et al. [60] have developed a scale for susceptibility to persuasive strategies. They focused on Cialdini's six influence strategies, which are reciprocation, scarcity, authority, commitment and consistency, consensus, and liking. They hypothesized that compliance to a request depends on the individual's susceptibility to persuasion, and confirmed this hypothesis in several studies [61, 63]. Later, they showed the personalization could be both explicit (i.e. measures based on questionnaires) and implicit (i.e. behavioral measures) [62].

As a conclusion of the studies, persuasive systems should be personalized, and system designers should measure their users' susceptibility to persuasion. Personalized persuasion strategies are shown to be effective upon health-related messages [61].

2.2.2.3 Anticipatory Mobile Computing

Given examples in previous sections evidently show how mobile phones change our behaviors and attitudes. In today's digital world, mobile phones are a very fundamental part of our lives. Smartphones are one of the greatest mediums for personalization and sending persuasive messages at the right time since they are personal, punctual, and suited for the user [64]. Recently, these capabilities of smartphones that affect human behavior have been proposed as "digital behavior change interventions (dBCIs)" [65]. Here is where the anticipatory mobile computing comes forward.

Rosen [66] defined an anticipatory system as "a system containing a predictive model of itself and/or its environment, which allows it to change a state at an instant in accord with the model's predictions pertaining to a later instant". Accordingly, anticipatory mobile computing can be described as the field, which is a combination of mobile sensing and machine learning to make intelligent reasoning according to the prediction of future incidents [64].

Mobile sensing is used commonly in health domain in order to detect patients' current situation and intervene them if needed. Mobile phones are actively used for prediction of behavior changes in diseases such as cold, fever, stress, anxiety, influenza, and mild depression [67]. One of the studies used smartphones to assess mental health and academic performance of students at Dartmouth College [68]. The activities and sleep patterns are easily collected with mobile phones, and these are correlated with self-reports so that mental health and academic performance are easily predicted. The trend goes to predicting communities' health situations with mobile sensor data such as the spread of epidemic diseases [69].

Almost all studies mentioned in Section 2.1 used *Experience Sampling Method (ESM)* as a methodology. ESM is used for capturing and recording human behaviors as it happens in their natural settings [70]. Hence, the data obtained by ESM has higher validity and less bias compared to other methods [71]. Since mobile technology has been developing day by day and has numerous context information as mentioned before, they have been used for ESM studies. Pejovic et al. [71] discussed the benefits and challenges of mobile ESM studies. They concluded their study by mentioning future directions of mobile ESM, such as integration with behavior interventions and anticipatory mobile ESM. These findings are also in parallel with anticipatory mobile computing.

Most of the anticipatory systems obtain data using ESM and the effectiveness of those systems mostly depends on user responses to the messages sent with ESM. Therefore, user responses become a prominent factor for a successful system. Responsiveness is commonly used as a synonym of receptiveness or attractiveness specifically in the mobile computing domain [72]. It is simply whether a mobile system user answers or reacts to the prompts generated by the system or not.

Several studies made an effort to infer responsiveness to mobile notifications using mobile phone related features such as application usage. Pielot et al. [73] stated that users are more open to receiving phone notifications if they have recently used their devices. Similarly, Mathur et al. [74] investigated the effects of several features such as the number of applications used in the last hour and the amount of time spent interacting with phone in the last hour on predicting user involvement with mobile phones. Their results show that involvement increases as the number of applications used in the last hour increases.

Responsiveness could be affected by personal factors such as the personality of users. The effects of the personality traits on responsiveness to mobile notifications have been investigated in a few studies. Mehrotra et al. [75] found that Extroversion trait had a significant effect. Yuan et al. [76] included personality traits in their models for predicting the responsiveness of users present in their training data set, and of new users, who were not present in the training data. Their method showed significant improvements when personality traits were included.

Responsiveness in the relevant literature has always been tackled in one dimension: whether a user responded to a survey or not. Although users may appear to be attentive to ESM surveys at first, their answers might be inaccurate, repetitive or random. Hence, response style of users should be understood in order to make more reliable inferences. Response style is defined as “a respondent’s tendency to responding systematically to questionnaire items regardless of the content” [24]. The most common response styles are acquiescence or disacquiescence (the tendency to agree or disagree to an item), extreme response style (the tendency to use the extreme categories), and middle response style (the tendency to use the middle category).

Recent studies showed that these response styles and personality traits are highly related even at the country-level [77, 78]. At individual level, the acquiescence response style was found related to the Extroversion [79], Openness, Conscientiousness [80], and Agreeableness [81]. Extreme response style was found to be related to Extroversion, Conscientiousness [80–83], Neuroticism [82], and Openness [80]. Finally, the middle response style was found related to Agreeableness in one study [81].

2.2.2.4 Mobile Interventions for Behavior Change

As previous sections present, mobile phones have become primary mediums for behavior change interventions in recent years. Mobile interventions in health domain have been mainly used for smoking cessation [84], improving the conditions of patients with diabetes [85], and promoting physical activity [86, 87]. The mediums that are used through mobile phones for health interventions are given as follows [88]:

- *Text messaging (SMS):* Interventions can be sent effortlessly since the nature of SMS is a push-based technology. Text messages are used especially for delivering reminders.
- *Cameras:* Cameras are mainly used for logging health-related behaviors, providing extra information for healthcare providers, and supporting self-management.
- *Applications:* Mobile applications are the mediums, which are used most widely for behavior change interventions. The well-known types of mobile applications:
  - Logging applications for diet, physical exercise etc.
  - Monitoring applications for keeping track of personal health information
  - Applications integrated with other devices (e.g. pedometers)
  - Games for teaching health-related behaviors
- *Sensors:* Sensors are widely used for recording health-related data for interventions.
- *Internet access:* The capability of internet access from nearly everywhere, enables uploading health-related data to providers' servers and thus, users can track, monitor their well-being easily.

Klasnja and Pratt [88] give five different intervention strategies used in mobile health domain: (1) Keeping track of health information (e.g. health-related behaviors, symptoms), (2) Involvement of healthcare providers (e.g. remote coaching, remote symptom monitoring, automated feedback), (3) Support of social influence (e.g. peer-to-peer influence, social support from friends and family, peer modeling), (4) Increasing access to health information (e.g. informational messages, reminders, easily visible displays), and (5) Using entertainment (e.g. games)

As stated above, reminders are a kind of "pushed" mobile interventions. The main aims for using reminders are increasing medical adherence and increasing health-related behaviors that individuals may forget to perform [88]. The design of reminders (e.g. content, frequency) is important for mobile interventions to be successful and effective. More specifically, reminders should be non-disruptive and motivational.

According to another review [89], the most common behavior change techniques in mobile environments are providing feedback on performance, goal setting, providing information on the outcomes of the behavior, tailoring, prompting self-monitoring of the behavior, and identifying barriers/problem solving/identifying ways of overcoming barriers.

2.2.2.5 Message Contents Used in Interventions

The design of message content used in reminders or any other intervention medium is one of the criteria that should be considered for an intervention to be successful or effective on users. Different strategies could be used for this purpose. For example; a reminder may be sent to users in order not to forget doing exercise or taking medicines. Another example may be a notification sent to a user's mobile phone after submitting an exercise log in order to reinforce the user for keeping him/her doing exercises. Several message content examples given in previous studies are summarized in Appendix A.2.

As can be seen from the table, the first example is not a mobile intervention; instead, the intervention was delivered through an e-health platform to diabetic subjects. Different cues were used in this study such as performance level and emotional status. The messages were delivered according to these categories. They were all encouraging and motivational messages.

The study of Fjeldsoe et al. [90] shows different message types depending on self-efficacy, outcome expectancy, goal setting, social support, and environmental opportunity. Pina et al. [91] show two different message types in their study: Positive reminder, and feedback. Positive and negative reminders are used for reminding users to take breaks, and feedback was delivered to users after completing a desired behavior (e.g. taking an adequate number of breaks in a day) so that users are encouraged to perform the same behavior in future. Several studies which are not mentioned in Appendix A.2 also use this technique in their interventions [92, 93].

Van Dantzig et al. [30] developed a persuasive mobile application for participants to take breaks after sitting for a long time. Instead, they delivered the messages through SMS. They compared the behaviors of the control and treatment groups. The control group did not receive any persuasive messages, whereas the treatment group received random persuasive messages which were formed depending on Cialdini's influence strategies [53]. The message types for authority, commitment, consensus, and scarcity are given in the Appendix A.2.

Tabak et al. [94] developed an algorithm, which produces real-time messages for Chronic Obstructive Pulmonary Disease patients. They separated message types into three: encouraging, discouraging and neutral. The reason for using discouraging messages is that patients having this disease should not be active too much in order to be protected from the symptoms (aches etc.). If the activity is at the desired level, then neutral messages were formed.

The studies show that appropriate messages should be sent at appropriate times. Message content is an important factor because an unrelated message to user's situation could be a failure for an intervention system. In addition to the findings of the studies which show the significance of intervention time, these studies show the importance of message content. Messages should be motivational and encouraging (most of the time) so that users should change their behaviors according to the desired target.

2.2.3 Behavioral Health Interventions for Work-Related Musculoskeletal Disorders and Sedentary Behavior in Office Environment

In this section, first, work-related musculoskeletal disorders are defined and behavioral interventions related to preventing them are presented. Then, sedentary behavior in office environment is discussed. Finally, the importance of self-awareness or self-regulation related to sedentary behavior is presented.

2.2.3.1 Work-Related Musculoskeletal Disorders (WRMSDs)

WRMSDs are defined as "impairments of body structures such as muscles, joints, tendons, ligaments, nerves, bones, or a localized blood circulation system caused primarily by the performance of work and by the effects of the immediate environment where the work is carried out" [14]. They are one of the most common work-related problems in Europe [14, 95]. Upper limb disorders are a type of WRMSDs, and they most likely affect hands, arms, shoulders, and neck [15]. The risk factors of upper limb disorders are repetition, postures while working, forces applied to these areas, and exposure duration.

Most of the office workers in today's world face repetitive movements and/or static postures in their work lives. Since we live in a digital age, most of the tasks are accomplished by computers. Computer use mostly affects upper parts of the body. Williams and Westmorland [96] stated that most common diseases related to computer use are carpal tunnel syndrome, tension neck syndrome, and thoracic outlet syndrome. Repetitive movements (such as using keyboard and mouse) result in muscle fatigue since muscles are not able to rest sufficient time, thus inflammation, degeneration and changes in tissues may occur. The patients with Repetitive Strain Injury (RSI) reported that their computer use is more than four hours a day [97]. Static postures while using computers result in muscle fatigue because of irregular blood flow in the body. Such WRMSDs result in unhealthy employees, reduction of efficiency and effectiveness, and economic costs for organizations.

Prevention from WRMSDs has been discussed widely in previous works. The organizations make prevention plans, interventions as well as Health and Safety Agencies throughout Europe. Health and Safety Laboratory in the United Kingdom published a report which explains the type of exercises that should be performed for preventing WRMSDs [15]. Additionally, the European Agency for Safety and Health at Work published a prevention report, which is a combination of workplace interventions, including behavioral modifications [14]. The most common characteristics of reports and research in this domain are that they recommend simple micro-breaks from repetitive works [12–15]. Henning et al. [10] suggested 30-60 seconds micro breaks from computer use at every 15 minutes in order to prevent musculoskeletal discomfort and increase the employees' productivity.

Similarly, McLean et al. [11] found that breaks at 20 minutes interval have a positive effect on reducing discomfort for computer terminal workers. Galinsky et al. [17] also showed the significance of breaks and their effects on eye strain and musculoskeletal discomfort. They also suggested stretching exercises for data entry workers during breaks. Friedrich et al. [16] also made suggestions for performing simple exercises at workstations so that the compliance could be improved.

2.2.3.2 Sedentary Work Style

Regardless of WRMSDs, sedentary working style of office employees causes many health-related problems. Warren et al. [9] declared that individuals with sedentary lifestyle have a higher risk of having cardiovascular disease. Similar to WRMSDs, taking breaks from sitting for prolonged periods can decrease the risk of metabolic diseases. For example; Healy et al. [8] showed the association between taking frequent breaks from sitting and a healthier metabolism.

In recent years, with the help of digitalized environments, there have been several studies for reminding office employees to take a break or make stretching exercises. One of the prior studies that was conducted in this area is the study of Monsey et al. [98]. They used a reminder software named "Stretch Break", and conducted study with treatment and control groups. All participants were informed about taking stretch breaks at 45 minutes interval reduces musculoskeletal discomfort and the risk of RSI. The treatment group used reminder software, whereas the control group did not. The results showed that the treatment group made stretching exercises more than the control group. Their results show that although the participants of their study were instructed to take breaks, most of the time they did not comply to take breaks. They concluded that there may be different factors that affect the decision of individuals for taking breaks. For example, an interesting point to be investigated is the relationship between musculoskeletal discomfort and compliance to the break-reminder programs/applications.

Trujillo and Zang [99] conducted a study in which they investigated the perception and satisfaction of data entry workers toward a stretching program named "Stop and Stretch". They have found that 63.3% of the workers stated that the program had a positive effect on their productivity. Another study [13] offered an interactive break reminder package named "Super Break", which encourages office workers to take more breaks. As a difference from previous work, Super Break detects keyboard and mouse activity between breaks and based on the activity it reminds taking a break or not. Additionally, it provides interactive games to users so that breaks can be more enjoyable and users are encouraged to take more breaks. According to the results of the study, 85% of the workers stated that they prefer Super Break compared to traditional break reminder programs. Similarly, Berque et al. [100] designed a software system, which persuades users to avoid immoderate typing speeds, to use typing shortcuts, and to take breaks from typing in order to prevent RSI. The system reminded participants when they exceeded a typing speed and gave feedback. The results showed that the feedback provided by the system had a positive effect on typing behavior, and shortcuts for words were used more effectively.

A more recent study [30] showed the importance of taking breaks and that it can be improved with persuasive strategies. They developed a mobile application named “Sit Coach”, which monitors the physical activity and reminds users to take breaks from sitting. The main aim of the mobile application is to decrease sedentary behavior of office workers. Users can configure the number of inactive minutes (default was 30 minutes), and at the end of this period “Sit Coach” sends a reminder. The experiment was conducted with treatment and control groups: The treatment group received persuasive messages about taking breaks whereas the control group did not receive any messages. The results showed that users who got intervention (got persuasive messages) reduced their computer use, and their physical activity increased compared to the control group.

Cooley and Pedersen [101] conducted a study with 46 participants for increasing non-purposeful movement breaks at work in order to reduce prolonged sitting times. They designed a persuasive software that reminds employees to take a break from their sitting times. They concluded that reminders should be unobtrusive for the system to be successful in the long term. As an opposite view, Wang et al. [102] stated that obtrusive reminders are necessary for non-difficult and easy-to-perform behaviors (e.g. brushing teeth, or taking breaks). In addition, doing simple exercises can be performed during coffee/tea breaks so that employees do not get overwhelmed.

2.2.3.3 Awareness of Sedentary Behavior

The number of studies in awareness about sitting behavior is limited in the literature. For example, in a qualitative study [103] related to work breaks, participants stated that they were not aware of how much time they have been working most of the time, and reminders from applications for taking breaks could improve their productivity. In another qualitative study, participants stated that lack of awareness related to physical activity affects it [104], or reversely, growing awareness could be a motivator for it [31]. van Dantzig et al. [30] also showed that the internal control toward sitting behavior was low for most of the participants of their study. Similarly, Wallmann-Sperlich et al. [105] stated that individuals who believe that sitting for long periods would not be harmful actually sit for a longer amount of time than individuals who do not. Those results show that personal beliefs and awareness regarding a specific behavior actually affect performing the behavior, and they show the importance of internal factors such as awareness regarding taking breaks. Luo et al. [26] recently explored the self-regulation and habit strength for preventing prolonged sitting via a mobile application and found that stronger self-regulation led to quicker responses to notifications. It is the only study that investigates the relation between self-regulation and responsiveness.

2.3 Work Engagement, Challenge, and Attentional States in Workplaces

Work engagement is referred as a state, which is active and positive, and it is described with a high level of energy, strong involvement to work, and a full concentration [106]. Challenge level can be described as the degree of the mental effort that should be exerted to complete a task [2]. In line with the study of Mark et al. [2], these labels are used as reference terms and they do not fully characterize the definitions in this study. More precisely, challenge level is used to specify user response of the question regarding how challenged a user is.

Boredom is described as "lack of stimulation or inability to be stimulated thereto" [107]. It comprises a penetrating deprivation of interest and difficulty of focusing on the ongoing task [108]. Individuals mostly seek a way to escape from the boredom state [109]. To date, several efforts have been made for predicting boredom. Physiological sensors [110] or logging computer activities [2, 23] are some examples of boredom detection techniques widely used in previous studies. With the increasing power of mobile devices, mobile device sensors have also been used for boredom detection.

Mark et al. [2] proposed a theoretical framework representing attentional states in the workplaces, given in Figure 1. They measured engagement and challenge levels of workers in workplaces via ESM questions, then separated the attentional states into four categories: (1) "*rote*" represents highly engaged, not challenged; (2) "*focus*" represents highly engaged and challenged; (3) "*bored*" represents low engagement, not challenged; and (4) "*frustrated*" represents low engagement, high challenge.

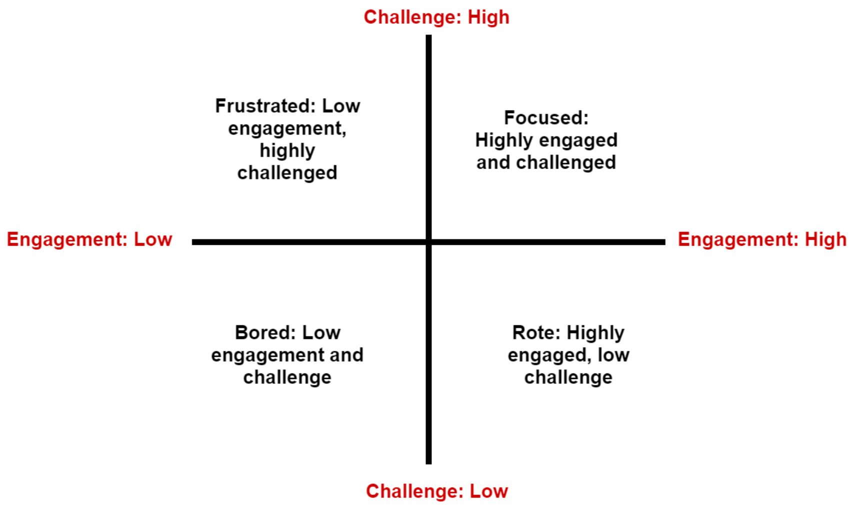


Figure 1: Theoretical framework of attentional states in workplaces (adapted from [2])

Numerous aspects of work engagement have been studied in the literature. Specifically, the relation between work engagement and personality has been widely discussed. Mark et al. [111] showed a significant positive relation between conscientiousness trait and engagement ratings. Bakker et al. [112] investigated the role of personality on the relation among work engagement, task performance and learning. They found that high conscientious employees had a higher level of work engagement. Similarly, Liao et al. [113] found that high extroverts, low neurotics and low conscientious employees have high work engagement levels. Another study [114] also showed a strong relation between personality traits and work engagement for the purpose of validating a work engagement scale.

Regarding work challenge, Mark et al. [2, 111] investigated the relation between challenge levels and personality traits, however, they did not find a significant relation. On the other hand, they presented a significant effect of Facebook use, e-mail use, and task switching between computer applications on both engagement and challenge levels of knowledge workers. According to their results, the duration of Facebook use is negatively related to engagement and challenge levels, whereas the duration of e-mail use and the number of application switches are positively related.

In another study [115], the authors investigated the relation between work productivity and application use, and found that the workers who use e-mail feel less productive. In addition, total computer application switches and the number of face-to-face interactions were also found negatively related to work productivity, i.e., the workers who perform higher switches between computer applications, or the ones who have higher number of face-to-face interactions in offices feel less productive. As contradictory to those results, Nduhura and Prieler [116] found that social media use and talking to friends during work hours make employees relaxed and energized, so that their work productivity increases.

Although it is not related to work engagement or challenge directly, it is worthwhile to mention a recent study [117], which developed a model for inferring low and high performers in workplaces through mobile sensing. They collected activity, location, phone usage (lock/unlock), light level through mobile application, heart rate and stress through a wearable device, and time spent in work and time spent at a break through a Bluetooth device. Their results show that higher performers unlock their phones less but they are more active than low performers. They built a XGBoost classifier for classifying the performers and the performance of the model is presented as AUROC=.83.

It was shown in a very recent study that focus level of users are effective on perception of health-intervention messages sent through mobile phones and adherence to the intervention messages [118]. Despite they did not limit their participants as office workers, the implications of their study show the significant effect of work engagement on the receptiveness of users.

Mark et al. [2] investigated which online activities are related to attentional states and how they are related in workplaces. Their results showed that type of online activity affects the attentional states of workers. For example; workers are usually in "bored" or "rote" states when viewing/writing e-mails, whereas they significantly spend less time in "focused" state when using Facebook or when surfing on the internet. In their another study [115], they investigated whether office workers felt an attentional state, then directed using Facebook, e-mail etc. Their results show that Facebook use after a "rote" state is significantly longer than Facebook use after "bored" state, and after "frustrated" state. In addition, workers use e-mail significantly longer when they feel "focused".

Another study [19] focused on predicting moods with wearable devices in work environments. They recorded heart rate, pulse rate, pulse wave transit time, accelerometer, and skin temperature. They fitted both personalized and generalized models upon the data obtained from the wearable devices for mood prediction, and their results show that personalized models for mood prediction are better than generalized models.

A number of studies worked on predicting boredom or general moods of mobile phone users. Even though they were not specifically designed for knowledge workers, their findings give insights related to which mobile phone features are effective on predicting moods. Pielot et al. [22] investigated which mobile phone features are indicative for detecting boredom. They stated that users are more likely to use higher number of applications when they are bored. Similarly, in another study [35], the recency of communication, intensity of phone usage, proximity and hour of the day were found related to detecting boredom, and sending proactive recommendations when boredom was sensed by mobile phones significantly attracted users' interests. Matic et al. [21] also found that the number of launched applications is a predictive feature for detecting boredom on smartphones. LiKamWa et al. [119] developed a mobile phone application, which predicted moods of users with smartphone usage patterns by fitting individual and general level classifiers. They found that phone calls and categorized application usage were the strong predictors for detecting mood. As mentioned above, the models presented in the studies predicted general mood of users, or boredom of general mobile phone users at any time.

2.4 Understanding Office Context

In one of the studies related to taking breaks and increasing physical activity during breaks [120], the authors suggested two main determinants leading to physical activity behavior: (1) attitudes, behavioral and social determinants, and (2) environmental and policy determinants. Knowledge, behavioral management skills, self-efficacy, enjoyment, perceived benefits, perceived barriers, and social support from family, coworkers and friends constitute the first category; whereas workplace norms and "culture", management support and available physical space constitute the latter. Hence, they emphasized the workplace routines have an important role for workers to take breaks or do exercises. Because of that, the social determinants in offices are explored in this study.

Subjective norm (social norm) is defined as "person's perception that most people who are important to him/her think he/she should or should not perform the behavior in question" [29]. It also plays an important role in the intention to use or adopt a new technology [121]. Hence, subjective norm is commonly used in the studies related to technology acceptance.

A recent study [122] was conducted in the office environment and focused on the effects of social norms on the adoption of mobile applications for promoting physical activity. They found that social influence is an effective factor for using such applications. Another study [30] showed that office employees are affected by their colleagues regarding their sitting behavior. Similarly, George et al. [31] emphasized the importance of social groups or being able to interact with others as a motivator for performing physical activity in a university. Hence, there is a need for investigating how office workers are affected by their peers or other environmental factors in offices for assessing responsiveness to health-related mobile notifications. To the best of our knowledge, there has been no study that investigates that direct relationship between subjective norm and responsiveness.

In another perspective, there are studies, which investigate the effects of office type on sedentary behavior [33] and distraction [123]. Mullane et al. [33] showed that the employees in private offices have a higher amount of sitting time compared to the ones in open offices, and they also discussed that employees in open offices might be more receptive to social cues than those in private offices. Seddigh et al. [123] showed the effects of personality and office type together, and concluded that Agreeableness and Openness traits are positively related to distraction, and the relationship is stronger among employees in open offices than the ones in cell offices. Finally, Morrison and Macky [32] showed that distraction is higher in shared offices. Hence, those studies give us an idea about office-related factors are important on perceived distraction (which might be considered as a reverse of work engagement), and also on responsiveness to rest break reminders.

2.5 Implications from Previous Studies

The first section has described the studies related to inferring opportune moments of users, and sending messages or interventions at those moments. It can be seen that context data obtained from mobile devices, desktop computers or wearable devices is important for identifying opportune moments of users. Most common features can be summarized as *time*, *application usage*, *ringer mode*, *physical activity* and *location*. Although there are studies which investigated interruptibility of office workers, as far as we know, there is no study related to rest break availability prediction. Investigation of such availability, which is a longer availability compared to "opportune moments" studies, is needed specifically for sending health interventions through mobile phones.

The second section, which gives details related to behavioral health interventions, has emphasized the aspect above. Specifically, the subsection related to WRMSDs has provided details regarding how those diseases could be prevented. Short breaks from working or sitting have been found as the most effective intervention in previous studies. There have been studies for reminding workers to take breaks in the literature and they can be “smarter” as mentioned before. In Section 2.2.2.3, responsiveness in "smart" environments has been discussed. In the studies related to anticipatory mobile computing, responsiveness has generally been measured as binary i.e., whether a user responded to a message or not. In the literature, there are metrics for measuring response styles such as middle response style or extreme response styles. As far as we know, those metrics have not been used on the responses obtained from an *in-situ* questionnaire. It is believed that assessing response style measures of office workers gives insights about their work environment, so that effective messages can be delivered to them.

The third section has presented the studies related to work engagement and challenge levels, attentional states of knowledge workers, and how boredom is detected through mobile phones. The relationship between work engagement/challenge and personality has been investigated in a few studies. Those studies measured general work engagement levels of employees for once with different scales/questionnaires. However, work engagement and challenge levels may change from time to time, or from task to task. It is needed to measure *in-situ* engagement and challenge levels, and to investigate the relationships with those measures. A few studies, which worked on *in-situ* engagement and challenge levels, investigated how those measures can be inferred with desktop applications. It may be interesting to study the relationship between mobile application usage and engagement/challenge levels.

Finally, the fourth section has mentioned the studies related to understanding office context. In this section, subjective norms in office settings, i.e. how office workers are influenced by their colleagues, are emphasized. The studies related to office type have been also presented. Those studies have showed that the influence of colleagues or office type have an effect on the behaviors of the office workers. In addition, the interruptions caused by colleagues or office type have an important effect on the engagement of office workers. Hence, subjective norm should be considered when investigating office workers.

CHAPTER 3

USER EXPERIMENT

In this chapter, the details of the user experiment are presented. In the first section, research framework and experiment design are given. Then, the mobile sensing application, UBDroid, is explained in detail. In the third section, the instruments used in the study are given. The pilot study and the participants are explained in the subsequent sections. Finally, data collection procedure is presented.

3.1 Research Method and Experiment Design

Implications from previous studies have been given in Section 2.5. On the basis of those studies, current study focused on the factors that affect knowledge workers for taking rest breaks, as well as *in-situ* engagement and challenge levels of knowledge workers. Hence, receptivity is extended in a way that knowledge workers respond to the messages about taking a rest break and his/her availability for the duration of that rest break (e.g. 10 or 15 minutes or more). In addition to that, several factors, that have not been investigated in the literature before, have been added to the research framework, which is developed for investigating engagement/challenge levels of knowledge workers. Hence, the effects of the new factors are presented as a result of this study. The frameworks for the sub-sections of the study are given in the related chapters (see Chapter 4 and Chapter 5).

Mobile-based Experience Sampling Method (ESM) was adapted throughout the study. ESM is used for capturing and recording human behaviors as it happens in their natural settings [70]. Hence, the data obtained by ESM has higher validity and less bias compared to other methods [71]. Despite some challenges (e.g. recruiting participants, sampling time, or technical challenges), ESM is a strong and powerful methodology for capturing users' natural feelings and thoughts.

The experiment designed for validating whether the factors are effective on the responsiveness of knowledge workers has four steps as depicted in Figure 2 [124]. As the first step, a questionnaire consisting of the following parts was applied to participants: (i) demographic information, (ii) information about routine break times, (iii) information about ringer mode of mobile phones, (iv) previous and current health status, (v) awareness about taking breaks, and pain or numbness level felt during computer use, (vi) stages of change scale, (vii) behavioral intention, (viii) social norms, and (ix) calendar use.

As the second step, mobile sensing application (UBDroid) was installed to participants' mobile phones. Participants were able to fill in two additional questionnaires namely Basic Personality Traits Inventory (BPTI) and Mobile Phone Problem Use Scale (MPPUS) via UBDroid whenever they wanted.

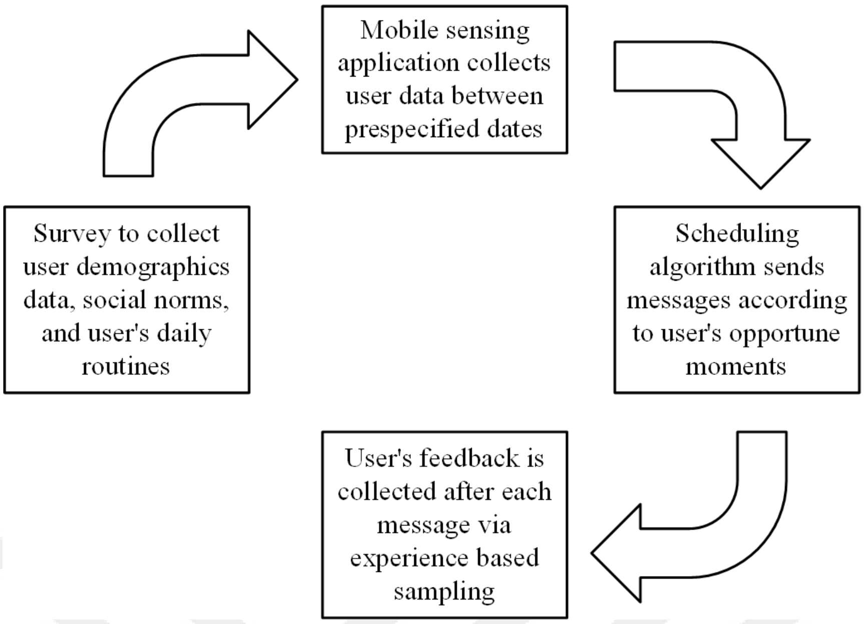


Figure 2: Experiment steps

The reminders were sent at the "opportune" moments that the scheduling algorithm chose for each participant. The opportune moments were identified based on the calendar events of the participants. The algorithm selected time slots only between the work start and end hours. In addition, it considered the routine break times stated in the questionnaire (Step I) by participants.

The reminder messages and the ESM questions were delivered to participants via UBDroid. These questions were related to what participants are doing when each reminder was sent, their challenge and engagement levels at that moment, their availability to take a break and to do an exercise at that moment. Hence, user feedback was collected after each reminder message. The content of the messages, and the motivation level at that day were evaluated by the participants at the end of each experiment day. The experiment lasted for 10 work days for the participants. The details of the experiment regarding the mobile sensing application (UBDroid), instruments used in the experiment, data collection procedure, and pilot study are given in the following sections.

3.2 Mobile Sensing Application: UBDroid

The mobile platform used as the main tool for data collection and notification delivery is UBDroid, which is an Android application developed by Akkurt [125, 126]. The application is designed for sending notifications to the participants, and gathering their responses and context data. Each device, which successfully installed the application, is recorded on the database with an anonymous unique identification number ("token"). Thus, the data of each participant will be kept and followed easily. The main functionalities of the application are given in the following section.

3.2.1 Capabilities

UBDroid has six main capabilities categorized under collecting mobile application usage, collecting context data, managing user groups, managing surveys and collecting responses, presenting videos, and message delivery.

- **Collecting mobile application usage:** UBDroid gathers application names that have already been installed on the phone, activity name that are currently being used on a mobile application, the time that user starts using specific application, and the time that user ends using that specific application. Mobile application usage is kept with "packet names" in logs. The screen on/off times and user's screen presence are captured at all times. Hence, it is possible to detect when a user starts using his/her mobile phone, which applications are used, and when the interaction with the phone ends.

- **Collecting context data:** In the background, UBDroid collects several context data namely Wi-Fi access point data, GPS location, accelerometer, activity type returned by the Google Activity Recognition API [127], ringer mode, and Google calendar data. The ringer mode is recorded every time it is changed. Activity data is only captured when accelerometer data indicates a change in the activity of a user such as walking or biking. The sampling starts whenever a participant starts moving and the activity type (e.g., walking, driving, and running) is stored in one-minute intervals until the user stops. In each state, transition from *still* to *moving* and *moving* to *still*, Wi-Fi access point data including BSSID, SSID, frequency and level information is recorded. For the devices that do not have significant motion sensor activity, accelerometer data is recorded every three seconds for one minute in five-minute intervals.

- **Managing user groups:** UBDroid system allows administrators create and manage user groups. This operation can be easily performed on the web interface of the system. As depicted in Figure 20 (see Appendix B.1), a group name and a description are required for adding a new user group. Then, registered devices (i.e. users) can be added to user groups as shown in Figure 21 (see Appendix B.1). User groups allow administrators to send a message to multiple users at the same time.

- **Managing surveys and collecting responses:** UBDroid allows administrators create new surveys on the system. Three types of questions can be added to a survey: comment, single selection, and multiple selection type. Comment type questions are used for open-ended questions. In single selection type of questions, users can only select one of the answers provided (such as Likert scale). A multiple selection type question allows users to select more than one of the answers provided (such as selecting interests). The web interface of UBDroid for creating a new survey is given in Figure 22 (see Appendix B.1).

- **Message delivery:** The main functionality of the application is delivering messages to the participants. The messages sent through UBDroid system are seen as a notification on mobile devices. The system allows sending messages to a specific user or to user groups. The surveys created before on the system can be added to messages. The web interface for sending a message to a user is given in Figure 23 (see Appendix B.1). The recipient of the message see the message as a notification on his/her mobile device as shown in Figure 24 (left). When s/he taps on notification, message can be seen as in Figure 24 (right). Messages can be sent to user groups as well. The web interface of UBDroid for sending message to a user group is depicted in Figure 25.

- **Videos:** A video menu named "Hareket Et" is added to UBDroid, so that participants can watch the exercise videos and do the exercises in breaks. There are two types of videos: short and long. Duration of the short video is about 5 minutes whereas the long one's is about 10 minutes. Participants may choose doing the exercises based on the duration of their availability. A sample screenshot of playing video in UBDroid is depicted in Figure 26 (see Appendix B.1). The exercises in the videos have been selected based on the literature and domain-experts.

3.2.2 Notification Delivery Algorithm

The notification delivery algorithm considers the duration between work start and end hours while sending the reminders to office workers. Digital calendar information is used in the algorithm to detect available moments and minimize the inconveniency. The algorithm receives the start and end hours of the events in digital calendars for identifying notification times. The notification times are identified in a way that they do not conflict with the event times. For example; if a participant is attending a meeting between 13:00-14:00, the algorithm excludes this time period from the duration between work start and end hours. In addition to the calendar event times, the algorithm considers the routine break times of the participants stated in the pre-experiment questionnaire. The timing of two notifications is selected from the routine break times. This is due to the fact that it is aimed to get a sufficient number of data points for modelling since the total duration of office workers' breaks are usually very short and not frequent. Then, the remaining four notification times are randomly picked from other available time period of participants (i.e. the time periods that the participant is not attending any event during work hours). If a participant has no preference, all six notification times are randomly chosen from the working hours.

The notification delivery algorithm for a user is depicted in detail in Appendix B.3. The steps can be explained as following (d presents the day number, i presents the notification number, p represents the preferred times of the user, e represents empty slots for current day, and r represents the random hour selected by the algorithm):

1. The algorithm starts at Day 1 ($d = 1$) with notification number 1 ($i = 1$).
2. The calendar events of the user are imported for day d.
3. The start and end times of calendar events are stored.
4. Empty slots for the current day e are calculated by extracting calendar events from user's work hours.
5. The empty slots are checked:
   (a) If there is not any empty slot for that day, that day is skipped. The algorithm starts the next day at Step 2.
   (b) If there are empty slots to send notification, then continue with Step 6.
6. The preferable times/hours of the user stated in the first questionnaire, p, are checked.
7. p can be an empty set because the participant might have skipped the question or left empty.

Table 2: Mobile devices used for testing UBDroid

Device Model	Android Version	Processor	RAM	Screen Resolution
Sony Xperia Arc	4.0.4	1 GHz Scorpion	512 MB	480 x 854
Turkcell Maxi Pro 5	4.0.3	1.4 GHz Qualcomm	512 MB	480 x 854
Samsung Galaxy Note 8	4.4.2	1.6 GHz Quad-core	2 GB	1280 x 800
LG G3	6.0	2.45 GHz Qualcomm	2 GB	1440 x 2560
LG G4c	6.0	1.2 GHz Qualcomm	1 GB	720 x 1280
LG G3	5.0	2.45 GHz Qualcomm	3 GB	1440 x 2560
General Mobile 4G	7.0	1.2 GHz Qualcomm	2 GB	720 x 1280
Samsung Note 4	6.0.1	1.9 GHz Quad-core	3 GB	1440 x 2560

(a) If p is empty (i.e. no preferred time/hour is found), then continue with Step 8.

(b) If p is not empty (i.e. preferred time/hour is found), it is checked whether the i is greater than 2 or not because maximum 2 notification times can be selected from p.

i. If $i>2$, then continue with Step 8.

ii. If if not, pick a random hour (r) from p. Then, delete r from p. It is checked whether r is a member of e:

A. If r is not a member of e, go back to Step 7.

B. If r is a member of e, then continue with Step 9.

8. A random hour/time (r) is selected from e.
9. ith notification time is set as r.
10. One hour before and after the notification time *[r-60,r+60]*, is deleted from e so that algorithm does not select those period as a notification time in previous steps.
11. The number of notifications (i) is increased as 1.
12. The number of notifications is checked whether it is reached to 6.

(a) If 6 notifications are not set yet, then go back to Step 6.

(b) If 6 notifications are set, then continue with Step 13.

13. The day number (d) is increased as 1.
14. The day number (d) is checked for whether 10 experiment days have passed or not.

(a) If d is less than or equal to 10, then go back to Step 2.

(b) If d is greater than 10, the experiment is ended.

3.2.3 Testing of the Application

Implementation and testing part include three different contexts: testing client-side of UBDroid, testing server-side of UBDroid, and testing notification scheduling algorithm. In order to test client and server-side of UBDroid the devices stated in Table 2 are used.

3.2.3.1 Tests on the Client Side of UBDroid

The client-side of UBDroid stands for the mobile application that the participants use during the experiment. Hence, in order the application to work properly during the experiment, testing the application and detecting the problems that might occur is important. Tests begin with the installation of the mobile application. A successful installation on Android 6 or above consists of the steps given in Appendix B.2.

1. In Step 1, the application asks for the permissions that it uses.
2. If user taps on “Install”, the application is installed on the device, however extra configurations are needed in order the application to run properly.
3. When user enters the application for the first time, it welcomes the user as given in Step 3.
4. After tapping on “Ileri” the application asks for the permission of application usage access. The permission is required for collection application usage data from the user. The user is informed as given in Step 4.
5. When user taps on “OK” button, the screen in Step 5 appears.
6. The user needs to allow usage access for BiMola application as given in Step 6.
7. Then, the user is directed to BiMola main page as in the Step 7.
8. The user taps on “Ileri” and it asks the username that will be used for the experiment. The user enters his/her username then taps on “Ileri” (Step 8).
9. The application read the device information, registers the device to Google Cloud Messaging service (GCM), synchronizes time, then completes the registration (Step 9).
10. Then, user is ready for using the application (Step 10).

The steps above are usually completed with success. However, there were cases that the application could not complete the registration because it could not register the device to GCM. These cases were mostly encountered with Turkcell Maxi Pro device.

After registration, the application should work properly, which means that it should not crush while collecting information. Sony Xperia Arc and Turkcell Maxi Pro devices could not handle this constaint while testing since they are older models than others. Their capabilities are not sufficient for the latest updates, so minimum Android version of UBDroid is set as Android 4.4.0.

Another parameter for client-side tests is correctly displaying messages sent by UBDroid. Specifically, the automatic messages sent via the notification delivery algorithm caused some messages to be unread and created a queue on the device. Hence, in some test cases, older messages were displayed to user, while newer messages should have been displayed and older ones should have been deleted. The user could not see the message in the message list, however, when a newer message arrived on the device’s notification bar, if the old one had not been read, then application displayed the older one. The problem was fixed in the latest version.

3.2.3.2 Tests on the Server Side of UBDRoid

Server-side tests includes two basic parts: controlling the database tables for accurate data flow, and controlling web-client for sending messages and commands.

1. When a new user is registered to the system, the *user* and *device* tables are controlled whether the user information is accurate.
2. *Controlling calendar events:*
   (a) When a new user registered to the system, the calendar events for past 30 days and next 15 days should be entered to the *calendar_event* table.
   (b) If a new event is added to the calendar or an existing event is deleted from the calendar for the current day, when "Get Calendar Events" command is sent to the user, the updates should be written on the *calendar_event* table.
3. *Controlling sensor information:*
   (a) If it is a weekday, the sensors start collecting data on 07:00 and end collecting on 19:00. The sensors collect data only if the device is in action (i.e. the user is walking, running, driving or tilting the device etc.). The data collected for a day is transferred to the server on the morning of the next day, or whenever the command from the web-client is sent. All sensor data except for Wi-Fi information is written to the *sensor_data* table, and Wi-Fi information is written to the *wifi_info* table. Hence, each weekday the *sensor_data* and *wifi_info* tables are checked whether the previous day's data has been arrived and written successfully. The time information that sensor data was captured (from 07:00 to 19:00), the types of sensors are all checked in this part.
   (b) If it is weekend, the sensors do not collect any information but only ringer mode of devices. Hence, no data should be written to the database on weekends. Again, the *sensor_data* and *wifi_info* tables are checked on weekends whether any information is written on weekends.
4. *Controlling application usage information:* The application usage information is collected all the time (not only for weekdays but also for weekends, and for all hours of a day). Hence, every day the *application_usage* table is controlled whether the application information is written correctly. The only exception for this control is that in Android 5 devices, the operating system does not allow user to select application usage access (Appendix B.2-Step 5 and 6) hence the application information is captured as only "Screen On" and "Screen Off". Since the analyses with application usage are not one of the main aims of the study, it can be acceptable as it is. In addition, when the number of devices with Android 5 is considered, they are expected to be minority in whole sample since most of the devices today use Android 6 or above. Because of these reasons, the application information from Android 5 devices can be ignored for the study.
5. *Controls for the web-client:*
   (a) A new message can be sent to users from the web-client. It is controlled whether the message is delivered to the correct users after sending it. If the message includes a survey, the message on the client-side is also controlled for displaying the survey correctly.
   (b) Commands can be sent for data transfer from the web-client.

i. *Transfer Collected Data:* It transfers all the sensor and application usage data that have been collected but have not been sent to the server. The command may not be delivered to devices in some cases because of network problems, hence, the control for the command is handled as explained in 3 and 4.

ii. *Get Calendar Events:* The command is used for updating the current day's events for the notification delivery algoritm. The control is handled as explained in 2.b.

3.2.3.3 Tests of the Notification Delivery Algorithm

Notification delivery algorithm runs depending on the working hours and event times of the user. It discards the time slots that the user has an event from the working hours, then if stated, it selects two notification times from the preferred time slots of users, otherwise it randomly generates the notification times. So, the notification times should satisfy the following constraints for each user:

1. The notification times should be between work start time and work end time. If these parameters have not been stated by the user, the algorithm automatically selects 7:00 as the work start time and 19:00 as the work end time.

2. The notification times should not overlap with the events. Every event between work start and end hours should be deleted so that the algorithm cannot select time periods that overlap with events.

3. If stated, two of the notification times should be selected from the preferred time slots of the user.

Tests for the notification delivery algorithm starts with fetching daily calendar events of users from the web-client. It updates *calendar_event* table so that if any event is added or deleted after registration, the algorithm does not miss the update. After running the algorithm, every user's notification times are checked with the event times of that day. There were minor problems in the algorithm causing overlaps with event times. These problems were fixed and the algorithm ran without problem.

The running of the algorithm is explained with an example. One of the users has working hours between 09:00 and 17:00. In Figure 3 below, the calendar events of that day are given. Based on the event times and work hours of the user, selected notification times for that day are shown in Figure 4. Note that notification times do not overlap with event hours and they are selected among work hours.

id	user_id	user_calendar_id	event_name	event_start_time	event_end_time
4800	22	51	-156765296	2016-12-12 08:30	2016-12-12 09:30
4804	22	51	1223823968	2016-12-12 17:00	2016-12-12 18:00
4805	22	51	69944179	2016-12-12 13:40	2016-12-12 15:40
NULL	NULL	NULL	NULL	NULL	NULL

Figure 3: Database table for the calendar events of a user

id	notification_time	user_id	survey_id	text	title
170	2016-12-12 16:10:00	22	21	biMola vakti!	Uzun süre oturmak modern dünyanın sigara alış...
171	2016-12-12 10:20:00	22	22	biMola vakti!	En iyi dinlenme anı dinlenmeye vakit bulamadığın...
172	2016-12-12 12:20:00	22	23	biMola vakti!	Uyarı: Aşırı yüklenme tespit edildi!!! Lütfen bir m...
173	2016-12-12 11:20:00	22	24	biMola vakti!	Bir mola ver ve içindeki sesi dinle.

Figure 4: Database table for the notification times of a user

3.3 Instruments

In this study, five instruments were employed for assessing different factors described in the research method part. The development and the details of the instruments are given in subsections.

3.3.1 Pre-Experiment Questionnaire

In total, 40 questions are included in the questionnaire given in Appendix C.1. The first two questions are related to demographic information: age and gender. Then, the questions related to occupation, company, position in the company, and duration in the position. In order to obtain work routines and regular work breaks, participants are requested to state their work start and end hours, regular break times, preferable times for exercises (i.e. the times that they want to receive reminders), why and how frequently they give breaks at work environments (e.g. coffee, smoking, chatting, meeting). The ringer mode of their mobile phone in regular break types and in regular daily cases (e.g. in theaters, restaurant, or with their friends) is also collected. The questions about health information about ergonomic problems consist of whether they have been diagnosed with WRMSD, and whether they have had treatment, and whether the treatment has been ended or not. This information is gathered because diagnosed patients might have a higher tendency to do exercises than healthy participants. Finally, the questions about calendar use of participants are included. Participants are requested to indicate how frequently they update their calendars, which events they include in their calendars, and how important to use a calendar.

In Question 28, the stages of change scale [44] is presented. In order to identify which participant is at which stage before the study, the scale is included in the questionnaire.

In Question 27, behavioral intention to give regular rest breaks and to use break reminder applications is measured. The questions are adapted from [128, 129]. In Question 31, social norms regarding giving regular rest breaks and using break reminder applications are assessed. Those questions are adapted from [128, 130]. The questions about behavioral intention and social norms are added to the survey since knowledge workers might be influenced from their roommates or colleagues to take breaks or to use reminder applications.

3.3.2 Basic Personality Traits Inventory (BPTI)

Basic Personality Traits Inventory (BPTI) was delivered to the participants during the experiment through UBDroid. The scale was developed by Gencoz and Oncul [131], specifically adapted and validated for Turkish people, includes the adjectives about personality characteristics. It was built upon the Five Factor Model of Personality, which consists of five personality traits (known as "Big Five"). These traits are Extroversion (i.e. "an energetic approach to the social and material world" [132]), Agreeableness (i.e. good-naturedness, cooperativeness, trustfulness), Conscientiousness (i.e. facilitating task- and goal-directed behavior), Neuroticism (i.e. being worrying, insecure, self-conscious, temperamental [133]), and Openness (i.e. originality, imaginativeness, intellectuality). A sixth factor (trait) named *Negative Valence* has been added in the study [131]. Negative Valence is described by being evil, awful, and cruel [134].

BPTI consists of 45 items with the internal consistency coefficients (i.e., Cronbach's alpha coefficients) of .89, .85, .85, .83, .80, .71 for Extroversion, Conscientiousness, Agreeableness, Neuroticism, Openness to Experience, and Negative Valence respectively. In this study, they were found as .90, .86, .90, .76, .74, .76 respectively for each trait. Test-retest reliability coefficients of the factors were reported as .84, .71, .80, .81, .83, and .72 respectively. The factor analysis showed the construct validity of the scale with the factor loadings varying between .81 and .63, .77 and .49, .68 and .53, .84 and .44, .70 and .63, .66 and .44 respectively for each factor. The scale is given in Appendix C.2. The BPTI scores of 19 participants are given in Appendix C.3.

3.3.3 Mobile Phone Problem Use Scale (MPPUS)

Mobile phones have become a part of daily lives, so it might cause problematic uses such as behavioral addiction. Bianchi and Phillips [135] developed the "Mobile Phone Problem Use Scale (MPPUS)" in order to measure excessive and/or problematic uses. The scale consists of 27 items that cover the dimensions of tolerance, withdrawal, craving, escape from other problems, and negative life consequences related to social, familial, work, and financial problems. The scale was selected since it has high internal reliability (Cronbach's alpha is found as .93 in the original study and it is found as .94 in this study). It shows moderate to high correlations with time spent using mobile phone ($r=.45$, $p<.01$), and established scale for measuring addiction ($r=.34$, $p<.01$) which supports the construct validity of the scale. In this study, the scale was delivered to participants during the experiment as a fun quiz, and not all the participants were required to fill in. It was not delivered with the pre-experiment questionnaire before the experiment began because it was not desirable to bother or exhaust participants with so many questions just before the experiment. So it was planned to deliver while the experiment continues so that the participants were able to fill it separately and without feeling an obligation. The scale was presented to the participants on UBDroid, so that they could fill in it whenever they wanted. The scores of the participants who filled the scale were calculated, and delivered them with a notification on UBDroid so that they could learn their degree of problematic phone use. The scale is given in Appendix C.4.

3.3.4 ESM Questions

The first and second questions were adapted from the study of Mark et al. [2]. Since it was aimed to measure *in-situ* engagement with work and challenge levels in order to investigate their relationships with other variables, the level of these measures were recorded each time message is sent. The third question was related to identifying whether participant was available for giving a break. If s/he was already in a break, the duration of this break was aimed to be recorded. The fourth question was related to identifying whether the break given or to be given is appropriate for doing simple exercises. Finally, the last question was for recording participants' behavior when the message was sent. The questions are given in Appendix C.5.

3.3.5 Post-Questionnaire

This questionnaire was adapted from the study of Koivumäki et al. [136], and it was delivered to participants after the experiment was conducted. The main aim of this questionnaire was assessing the satisfaction levels of participants through the mobile application used in the study, and the usefulness and efficiency of the application. Two open-ended questions were included in the questionnaire in order to record the feedbacks, comments, feelings of the participants. The questionnaire is depicted in Appendix C.6.

3.4 Reminder Messages

Each time a reminder was sent, a motivational message for taking a break or doing a simple exercise was sent to the participants. In total, 60 motivational messages were identified (6 reminder messages for 10 experiment days), and given in Appendix C.8. The messages were adapted/translated from the previous studies in the literature.

3.5 Pilot Study

The real experimental phase was initially evaluated with a pilot study, and the problems that might occur during the experiment were identified and solved. All the real experiment phases were tested with five participants having different mobile phones in terms of brands and Android versions and working in different workplaces. All the phones had significant motion sensor. In this phase, both accelerometer data and activity data were collected to be used later for training purposes since there might be users having mobile phones with no significant motion sensor in the real experiment. The participants were requested to fill a form indicating their location information at the end of each experiment day. To facilitate the process, all the locations where each user stayed for more than five minutes were automatically identified, and presented them with their timestamp. Finally, the participants labelled them. A sample form is given in Figure 5.

Başlangıç Saati	Bitiş Saati	Ofis	Ofis dışı iç mekan (indoor) bir yer	Dışarı (açık alan/outdoor)	MekanNo
27.2.2017 06:38:20	27.2.2017 09:25:35		x		1
27.2.2017 10:11:58	27.2.2017 10:52:20	x			2
27.2.2017 10:59:11	27.2.2017 12:05:30	x			2
27.2.2017 12:20:03	27.2.2017 12:37:01			x	3
27.2.2017 12:46:39	27.2.2017 15:13:59	x			2
27.2.2017 15:20:29	27.2.2017 17:02:39	x			2

Figure 5: A sample location form that pilot study participants filled each study day

3.6 Participants

The population of the study was selected as knowledge workers because they have prolonged sitting times in front of computers, and they have a sedentary lifestyle because of their working style. Such a lifestyle and prolonged sitting times cause WRMSDs (see Section 2.2.3). In order to prevent such diseases, regular rest breaks and doing simple shoulder, neck and wrist exercises are suggested. However, knowledge workers such as programmers need to focus deeply on their tasks to work more efficiently, and this causes them to forget taking breaks.

In total, 55 attempts were made for responding to the pre-experiment survey, and 50 individuals responded in full. Forty-two of them successfully installed the mobile application. Eleven participants dropped out of the experiment, resulting in a total of 31 participants. In the analyses, a varying number of people's data was used due to the fact that not everyone fully participated in all steps of the experiment. As a result:

- Nineteen of the 31 participants had a response rate of 25% or higher in ESM questions.
- The application package names were obtained from 24 of the 31 participants.
- The number of participants who filled BPTI out of 31 participants was 19.

Table 3: Descriptive statistics of the participants ($N = 31$)

Gender	Female	20 (64.52%)
	Male	11 (35.48%)
Age	Average	31.52
	Min	24
	Max	42
	Std. dev.	5.01
Occupation	Engineer	11 (35.48%)
	Academics	9 (29.03%)
	Specialist	6 (19.35%)
	Manager	4 (12.90%)
	Technical personnel	1 (3.23%)
Organization sector	Private	14 (45.16%)
	Government	13 (41.94%)
	Freelance	3 (9.68%)
	Owner/partner	1 (3.23%)
Work duration (in hours)	Average	8.5
	Min	7
	Max	10
	Std. dev.	1.15

The descriptive statistics of the participants are given in Table 3. Twenty of the 31 participants (64.52%) were male, and 11 of them (35.48%) were female. The average age of the participants was 31.52, with a minimum of 24 and a maximum of 42. The average work duration per day was 8.5

hours, with a minimum of 7 and a maximum of 10 hours. The job titles of the participants were varied. Eleven of the participants (35.48%) were engineers, nine of them (29.03%) were academics, six of them (19.35%) were specialists, four of them (12.90%) were managers, and one of them (3.23%) was technical personnel. Fourteen participants (45.16%) work in the private sector, 13 of them (41.94%) work in the government sector, 3 of them (9.68%) work as freelancers, and one participant (3.23%) was the owner/partner of a company.

3.7 Data Collection Procedure

Before conducting the experiment, the permission from METU Research Center for Applied Ethics was obtained (see Appendix C.7). Participants were selected with convenience sampling method: potential participants (who are office workers) were invited to join the experiment through several channels. The experiment was announced among the graduate students of the Informatics Institute in METU and promoted on social media. Over 90% of the students at the Institute work in a company or an organization (e.g. engineers, research assistants, specialists etc.). The leaflets that introduce the experiment were prepared and distributed. The participants were directed to the website of the experiment. The pre-experiment questionnaire and the mobile application download links were added to the website. Hence, participants were able to download the application just after filling the questionnaire.

The participants filled the pre-experiment questionnaire and installed UBDroid into their mobile devices. The experiment started on the day after they filled the questionnaire and installed the application. The whole data was collected between March 13 and April 10, 2017. For 10 workdays, a maximum of six reminder messages were delivered. The participants were granted a coffee cheque if they filled the pre-questionnaire and replied at least 25% of the ESM messages sent to them.

CHAPTER 4

A HYBRID MODEL TO PREDICT OFFICE WORKERS' AVAILABILITY FOR REST BREAKS USING MOBILE SENSORS

This chapter mainly explains the hybrid model built upon the data collected with the user experiment. In the first section, a brief introduction why the study was conducted is presented. Then, the methods used in the hybrid model are explained in the second section. Then, the methodology is presented. The details of the hybrid model are given with the parameter setting and feature extraction. The efficiency of the hybrid model is presented by comparing with other methods commonly used in the literature. The chapter ends with the results and the discussion of the factors, which affect office workers' break availability for rest breaks.

4.1 Introduction

It has been widely investigated in which situations users respond to mobile notifications and those notifications do not interrupt or bother users, and the context information that identifies those situations in several studies. Based on the findings of those studies; time, users' activity, location, mobile application usage, cognitive context or ringer mode of mobile phones are effective on the decision of responding to a notification or not [1, 75, 137–144]. Recently, studies for automatically identifying office workers' interruptibility using biometric sensors or mobile phones have been proposed [18]. In the recent interruptibility studies, the use of individual models has an increasing trend. Surely, individual models offer personalized solutions and increase the efficiency of mobile systems. However, the cold-start problem, which is the lack of sufficient training data points at the beginning of model development, has been seen as a major problem for building individual models. In such situations, the studies offer using generalized models, which are built upon all users' data and have higher number of data points, at the beginning, and switching to individual models after sufficient data points are obtained from each user.

Besides the interruptibility of office workers, there have been mobile systems, which offer taking rest breaks in order to decrease the sedentary work style of office workers. It has been known that taking a micro break at least once in an hour prevents several work-related musculoskeletal disorders [12–15]. Those systems suggesting to take rest breaks mostly focus on the degree of promoting active moments or taking breaks. As a future work, they state that the systems should be more intelligent to detect the moments for rest breaks. For this reason, a comprehensive study, which investigates the moments that office workers tend to have rest breaks, is needed.

In the study presented in this chapter, considering the two main gaps in the literature, a hybrid model is developed for investigating the effective mobile context data on taking rest breaks, and as a solution to the cold-start problem. Mainly, the study is targeted to find a solution to the first research question of the study: *"How can a model be built for inferring availability of office workers for having rest breaks using mobile phone sensors by considering cold start problem, the variety in the number and characteristics of the responses, and repeated-measures design of the data? How is this model comparable to individual and general models?"*.

4.2 Background

In this section, the details of the methods used in the hybrid model are given. First, Generalized Linear Mixed Methods are explained. Then, kernel density estimation is presented. Finally, repeated-measures correlation is defined.

4.2.1 Generalized Linear Mixed Models (GLMM)

Generalized linear models (GLM) extends linear models by handling response variables with non-normal distribution [145, 146]. GLMMs incorporates random effects to GLMs. Random effects are mostly individuals, population, species, or vials with lots of levels [145]. In its simplest form, a GLMM can be written as in Equation 1 where x is the vector of fixed predictors, and z is the vector related to random predictors. Fixed and random predictors have related parameter vectors β and b, and β_0 is the residuals vector.

$$y = \beta_0 + x\beta + zb \tag{1}$$

GLMMs can deal with numerous response distributions and repeated-measures observations. Specifically, the use of GLMMs is appropriate when there are many levels (e.g., individuals, species), few data on each level, or when the number of samples at each level is not the same. Although GLMMs are fast and powerful, with non-Gaussian response variables the likelihood cannot be obtained in the closed form [147]. A Bayesian framework using Markov chain Monte Carlo (MCMC) methods is a convenient way to fit a GLMM [145], which has recently become popular in the areas where several species or populations occur such as ecology, biology or zoology [148, 149].

Deviance Information Criterion (DIC) is generally used for model selection, and it shows the performance of a Bayesian model based on the deviance and the number of parameters by emphasizing the random effects. Lower DIC values should be preferred for model selection [150]. Furthermore, the convergence of the Markov chains is checked by the Gelman-Rubin diagnostic criterion where 1.002 and below indicates the convergence [151]. Finally, autocorrelation between consecutive iterations in the chain should be less than .10, which indicates the chain has mixed well.

In this study, the data points could not be considered as independent since the data set consisted of multiple responses of participants. Besides, the number of responses is not equal for each participant. Moreover, the response variable is ordinal, which means that its distribution is not Gaussian. For these reasons, the assumptions of approaches such as ANOVA have been violated. In this case, it is suggested to use approaches such as generalized linear mixed models (i.e., hierarchical modelling or multilevel modelling) [37, 152]. Because of the reasons above, GLMM has been adopted in the modelling phase, and the details are given in next sections.

4.2.2 Kernel Density Estimation (KDE)

Density estimation is basically reforming the probability density function using a set of given data points. Histograms are the most basic forms of the density estimation. A histogram needs a bin width and a starting point of the first bin. Hence, the starting point of the first bin, and the number of bins affect the density estimation. Besides, the density estimation obtained from histograms is not smooth. Because of those, histograms are inappropriate for most of the practical work.

On the other hand, kernel density estimation (or estimators) (KDE) are commonly used. A kernel function is applied for each data point in KDE. Using a smooth kernel function gives a smooth density estimate. So, the drawbacks of the histograms have been removed. More formally, KDE takes weighted local density estimates at each observed data point (x_i), then, aggregates them to derive an overall density.

The definition of the kernel density estimation can be given in Equation 2, where $K(x)$ is the kernel function and $h > 0$ is the smoothing bandwidth. h controls the size of the neighborhood around, which means that it controls the smoothing. A very smooth density is obtained with a large bandwidth, where an unsmooth (or with a high-varience) density is obtained with a small bandwidth. The kernel (K) controls the weight of the data points (X_i). A smooth unimodal function, which has a peak at zero, is generally used as the kernel function. Some examples of the kernel functions can be given as Gaussian, Epanechnikov, uniform, cosinus, and tri-cube.

$$\hat{p}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \tag{2}$$

4.2.3 Repeated-measures Correlation (rmcorr)

Repeated measures correlation is presented as a statistical method in order to identify the within-individual relationship for paired measures [37]. It basically refers to the basic correlation or regression techniques (e.g. Pearson correlation), which require independency between observations [153]. Those simple techniques are mostly used on aggregated data or data with non-independent observations (for example; nested data consist of multiple observations, which are collected from each participant [152]), but it causes erroneous or biased results, falsified type I error, or affects the statistical power [37, 152]. As an alternative, Bland and Altman [154, 155] offered calculating the within-participants correlation, which is referred as the repeated-measures correlation (rmcorr) by Bakdash and Marusich [37], which does not violate the assumption of statistical independence, and it has a higher statistical power.

According to the study [37], rmcorr presents each individual's linear fits with regression lines with the same slope but with different intercepts. The coefficient of rmcorr is between -1 and 1 similar to Pearson correlation coefficient. Rmcorr considers non-independence among data points, hence, it has a higher statistical power than the data aggregated for satisfying the Independent and Identically

Distributed (IID) assumption in simple correlation. Rmcorr can be considered as a multilevel model since it fits different intercept but a single slope for each individual in the model. The benefits of the multilevel models come from this point: Individual differences can be investigated as well as overall differences.

4.3 Method

In this first study of the thesis, a hybrid model is proposed for analyzing availability of office workers for taking rest breaks during work hours with mobile phone sensors. The data is obtained from the user experiment explained in the previous chapter. The data of 19 participants whose response rate is higher than 25% was used. The first three questions and the fifth question of the ESM questionnaire were included in the analyses of the first study. The ESM questions used are given as follows:

- **Question 3:** How long is your current break duration or for how long are you able to take a break? **Options:** Cannot take a break now, less than 5 minutes, between 5-15 minutes, between 15-30 minutes, and more than 30 minutes

- **Question 5:** What are you doing now? **Options:** In a meeting, working on computer, in a tea/coffee break, in a lunch/snack break, on road, in a social break (chatting, talking on the phone etc.), in a bathroom break, in a smoking break, following the media (news, magazines etc.), in a health break, in a praying break, and other

Nineteen participants responded to 528 ESM questionnaires in total. The number of responses in each category of break availability and break type is given in Figure 6. The number of responses marked with “15-30 minutes” in break availability was relatively low compared to other categories. Hence, the categories of “15-30 minutes” and “more than 30 minutes” were grouped together as “more than 15 minutes”.

In this study, break availability was modelled with the context data collected through mobile devices. As shown in following subsections, break type (which is assessed with the ESM Question 5) was not included in the modelling since break availability gives an idea about the break type. The research framework is given in Figure 7. Break availability was hypothesized to be affected by location, ringer mode, time, application usage and activity.

4.4 Training and Parameter Setting

Before continuing with the hybrid model, the training and the parameter setting phase is explained in this section. Specifically, the details of how features extracted from the mobile sensors are given. The data collected in the pilot study is used to determine the parameters, which is used later in the modelling phase of the final experiment data.

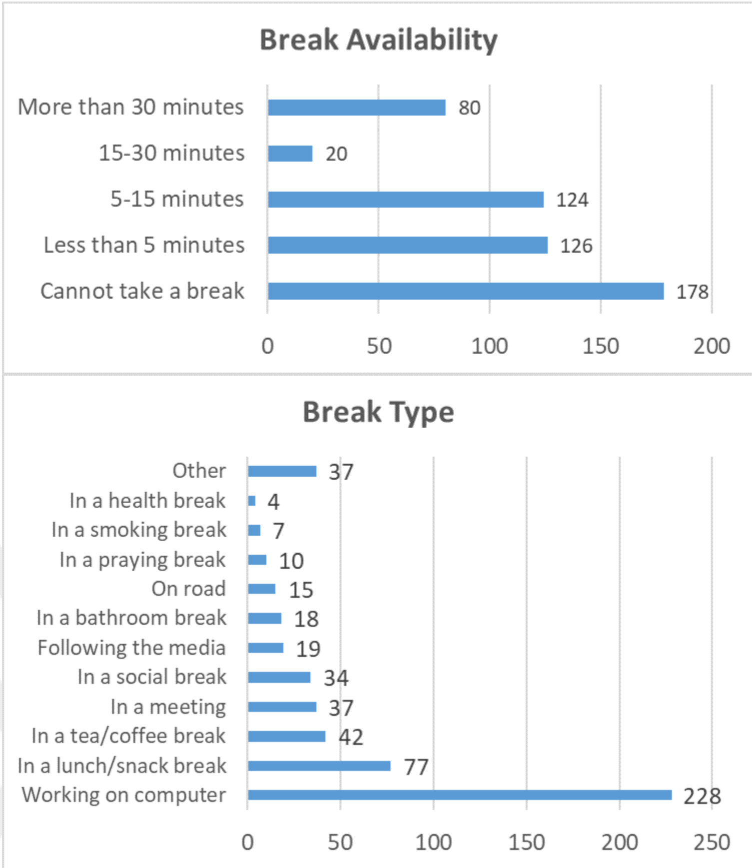


Figure 6: Number of answers in the categories of break availability and break type

4.4.1 Modelling User Activity

User activity is defined as whether a user is *moving* or *still* when notifications are sent. In order to identify user activity, *still periods*, i.e., the periods that the mobile phone is marked as being still for more than five minutes were extracted from mobile sensors. Figure 8 illustrates the timeline of a workday for a user. Every still period has a start t_{start_j} and an end time t_{end_j}, where *j* is the index for the *still period*. The start time of a *still period* is the moment at which the activity is read as *still* with an accuracy of 100% by the Google Activity library. The end time of a *still period* is the first moment the activity type returns a value other than *still*. In the figure, blue and green highlighted bars show the *still period* in locations l_1 and l_2.

4.4.2 Predicting Activity from Accelerometer Data

The smartphones of five participants did not show significant motion sensor activity, hence, the sequence of their activity types could not be directly recorded in the database. For this reason, a classification model was fitted in order to classify their activity types based on their accelerometer data. Ustev et al. [156] used three features for classifying activity data: standard deviation of magnitude of accelerometer data, variance of magnitude of accelerometer data, and mean z value of accelerometer

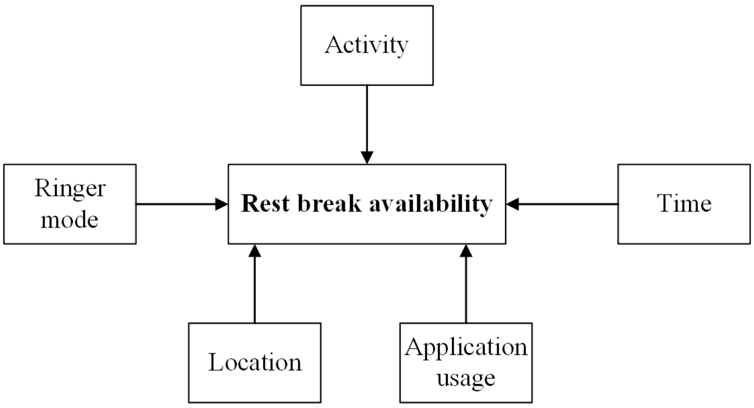


Figure 7: Research framework proposed in the first study

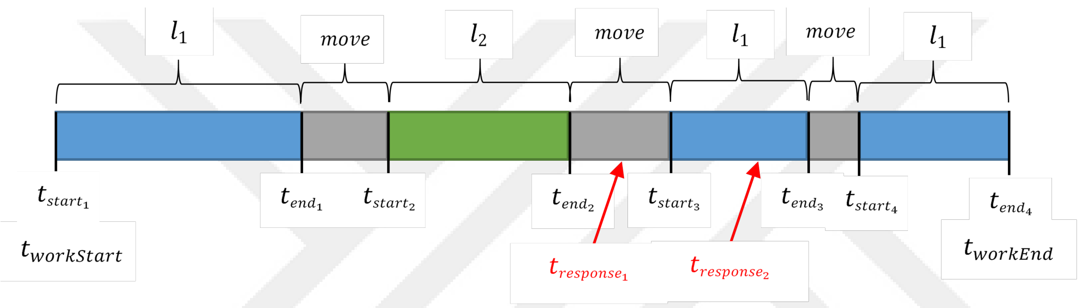


Figure 8: Timeline of a workday of a user

data. They used k-NN classifier and selected k as 50 [157]. Hence, k was selected as 51 since it is more appropriate to choose k as an odd number. The algorithm was trained and tested with the data obtained from smartphones with a significant sensor (i.e. the devices that the activity type is known) of the participants in pilot study. Then, two k-NN models were trained, one for actual activity types with six-level, (*still*, *in vehicle*, *on bicycle*, *on foot*, *running*, and *walking*), and one for binary output (for *still* activity type as 1, for *moving* activity type as 0). Since the activity types except *still*, indicate a moving activity, hence, they were aggregated as *moving*. In addition to k-NN classifiers, support vector machine (SVM) classifiers with radial basis kernel were developed on the data for actual activity types with six-level, and for binary output in order to compare the results with k-NN algorithm. The kernel width of SVM model was obtained through cross-validation. The accuracy results of each model are given in Table 4.

Table 4: Accuracy of different classifiers for activity classification

Model	**Output**	**Accuracy**
k-NN with k=51	6 activity types	75.90%
k-NN with k=51	2 activity types	81.40%
SVM	6 activity types	76.40%
SVM	2 activity types	82.43%

As can be seen from the table, SVM model with two activity types (*still* or *moving*) had the highest accuracy among all models. Hence, the classification of activity for the users without significant sensor was done with SVM giving binary output. After classifying the activity, *still time periods* could be computed for all users.

Equation 3 shows how the activity was labelled: $activity_n$ denotes the activity of the user when notification *n* is received. If the user's response time $t_{response_n}$ to notification *n* is between any start time and end time of still periods, then the activity is labelled as "*still*". Otherwise, it is labelled as "*moving*".

$$activity_n = \begin{cases} still, & \text{if } t_{start_j} \leq t_{response_n} \leq t_{end_j} \\ moving, & \text{otherwise} \end{cases} \quad \text{where } j \in SP. \tag{3}$$

4.4.3 Modelling User Location

Even though GPS data was recorded during the experiment, it is unreliable for indoor localization [158]. Since the participants of the study spend their most times indoor (in their offices), there is a need for differentiating user indoor locations so, RSSI fingerprinting based on Wi-Fi access points (APs) is used. Briefly, this method computes the similarity between two locations based on the RSSI of Wi-Fi APs recorded at those two locations. The similarity computation is given in Equation 4 adapted from [158]. Wi-Fi APs recorded at locations l_1 and l_2 are denoted as AP_1 and AP_2, $AP = AP_1 \cup AP_2$. $f_i(a)$ denotes the RSSI of AP, $a \in AP$, recorded at location l_i.

$$S = \frac{1}{|AP|} \sum_{\forall a \in AP} \frac{min(f_1(a), f_2(a))}{max(f_1(a), f_2(a))} \tag{4}$$

For each user, a similarity matrix with a dimension of *M*x*M* was computed where *M* is the number of still periods of that user. At first, each still period was considered as a single location. However, the user may be in the same location at different times of the day. In order to identify the locations of a user, hierarchical clustering was used since it enables to adopt a threshold value for identifying clusters. Different threshold values were attempted for identifying clusters. The accuracy of the method was computed by comparing cluster values with the real labelled locations of the participants in the pilot study. The total number of the labelled locations was 415.

Threshold values of .05, .10, .15, .20, .30, .40 and .50 were tested for clustering. The accuracy of each threshold is given in Table 5. The threshold value of .15 was selected since its accuracy is the highest among all values. The main reason for not achieving a higher accuracy value may be caused by user statements in the pilot study. The pilot study participants filled the location form provided to them, however, they may have forgotten the locations where they had been for a very short period of time, so that they may have specified wrong location in the form for that time period. Since the data obtained from the participants gave the most accurate result when the threshold was set to .15, the two still time periods, which had a similarity higher than .15, were considered as the same location. Still, remembering this shortcoming of the clustering, all threshold values were also given to the models as described in next sections.

Table 5: Accuracy of different threshold values for hierarchical clustering

Threshold	**Accuracy**
.05	62.70%
.10	61.90%
.15	63.10%
.20	61.90%
.30	57.10%
.40	55.40%
.50	53.70%

An example of the output of the clustering algorithm is given in Figure 9. The first column named "start time" shows the start time of a still time period, and the second column named "end time" shows the end time of the time period. For example; that user did not move with his/her smartphone from 09:00:21 to 10:10:21. Then, s/he was on move until 10:25:03. "Location" column shows the clustering output. For example; that user was in the same location (Location 1) from 09:00:21 to 12:11:36 even though s/he moved. Then, from 12:16:22 to 12:46:21 the user was in another location (Location 2).

	startTime	endTime	location
1	2017-03-28 09:00:21	2017-03-28 10:10:21	1
2	2017-03-28 10:25:03	2017-03-28 10:35:03	1
3	2017-03-28 10:41:04	2017-03-28 10:57:21	1
4	2017-03-28 11:02:34	2017-03-28 12:11:36	1
5	2017-03-28 12:16:22	2017-03-28 12:46:21	2
6	2017-03-28 13:05:03	2017-03-28 14:26:28	1
7	2017-03-28 14:40:34	2017-03-28 14:55:21	1
8	2017-03-28 15:05:03	2017-03-28 15:20:31	1
9	2017-03-28 15:33:22	2017-03-28 15:44:16	1
10	2017-03-28 15:53:21	2017-03-28 16:23:32	1
11	2017-03-28 16:30:21	2017-03-28 17:04:27	1
12	2017-03-28 17:10:16	2017-03-28 18:15:21	1
13	2017-03-28 18:30:21	2017-03-28 18:40:21	3

Figure 9: An example of clustering output for a user

4.5 Feature Extraction

The feature space is defined before modelling the data. Since the exact locations of users are not detected, new features are defined for representing locations in terms of duration, similarity, and frequency using the location clusters.

4.5.1 Location Features

Four features were introduced regarding the locations of the users. Those features are time spent in location, base location, location similarity, and location frequency. The details of the features are given in the following sub-sections respectively.

4.5.1.1 Time Spent in Location (TSL)

After defining user locations, the locations where users answered ESM message was extracted. Time spent in location means the duration (in minutes) spent in that location until the moment that ESM message is answered. The duration calculation is as following:

$$TSL_n = \begin{cases} t_{response_n} - t_{start_j}, & \text{if } activity_n = \text{"}still\text{"} \\ t_{response_n} - t_{start_{j-1}}, & \text{if } activity_n = \text{"}moving\text{"} \end{cases} \quad \text{where } j \in SP \tag{5}$$

In Figure 8 the calculation for location duration is illustrated. Assume that the user responded the first notification at $t_{response_1}$ while moving. Then, the same user responded to the second notification at $t_{response_2}$ at location 1 (l_1). The time spent in location for the first notification (TSL_1) is equal to $t_{response_1} - t_{end_2}$ since the notification arrived when the activity was labelled as moving (not in a specific location where a rest break is spent) and t_{end_2} is the time when the user left the pre-location (location 2). The difference between $t_{response_1}$ and t_{end_2} indicates the total time spent while the user was moving. The time spent in location for the second notification (TSL_2) is equal to $t_{response_1} - t_{start_3}$ because this time notification arrived when the user is at location 1 and t_{start_3} is the time when the user arrived at location 1. The variable is normalized by taking its natural logarithm.

4.5.1.2 Base Location

Since the experiment was conducted during the work hours of participants, it is assumed that the location each participant spent their time at most is their work places. Based on the duration values spent at each location, the location with the highest duration was selected as the user's work place.

4.5.1.3 Location Similarity (LS)

Location similarity is defined as the similarity between the location at the time the ESM message arrived and the user's base location. If the user is in the base location when the message has arrived, then the location similarity is set to one. The location similarity is obtained from the similarity matrix explained above.

4.5.1.4 Location Frequency (LF)

Location frequency is defined as the total number of visits to the location *i* throughout the experiment over the total number of visits to all locations throughout the experiment. For example, while the frequency of visits to the base location is expected to be the highest, the frequency of the locations where participants have their lunch is expected to be significantly lower, such as 10 visits if the individual had lunch every day at the same place during the experiment.

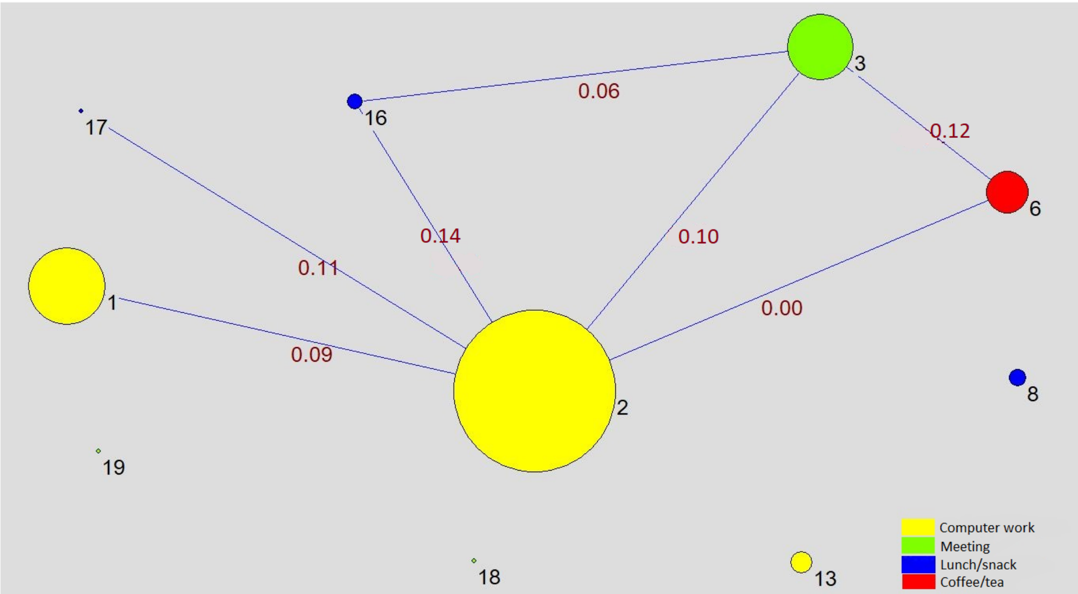


Figure 10: Location graph of a user

Figure 10 shows the location graph of a user. Each node in the graph represents a location of the user. The lines between the nodes represent the similarity of those two locations. For example; the similarity between location 1 and 2 is equal to .09. The size of the nodes represents the total duration spent in that location. The colors of the nodes are extracted based on the ESM responses which is related to the break type (Question 5 in ESM questionnaire). Note that the biggest node is the work place of the user (in this graph location 2). The second biggest nodes are the location 1 where the user works with his/her computer, and location 3 where the user attends his/her meetings. Location 8 is a totally different place compared to user's workplace because there is no connected line between Location 8 and Location 2. Location 8 might be a cafeteria where the user had his/her lunch.

4.5.2 Ringer Mode Features

In the pre-experiment questionnaire, it is collected in which ringer mode participants keep their mobile phones during their office hours and breaks. Figure 11 shows the responses of the participants. It can be seen that users keep their phones in different modes in different situations. In praying, health breaks, and in meetings, the participants stated that they keep their mobile phones in either silent mode or in vibrate mode but not in sound mode. Another important finding is that users set their ringer mode differently according to their office norms.

Out of the 528 ESM messages, 337 messages arrived when the mobile phone was set to normal (sound is enabled) mode, 180 messages arrived in vibrate mode and 11 of them at silent mode. Since the number of data points for the silent mode is significantly lower than the others, silent and vibrate mode

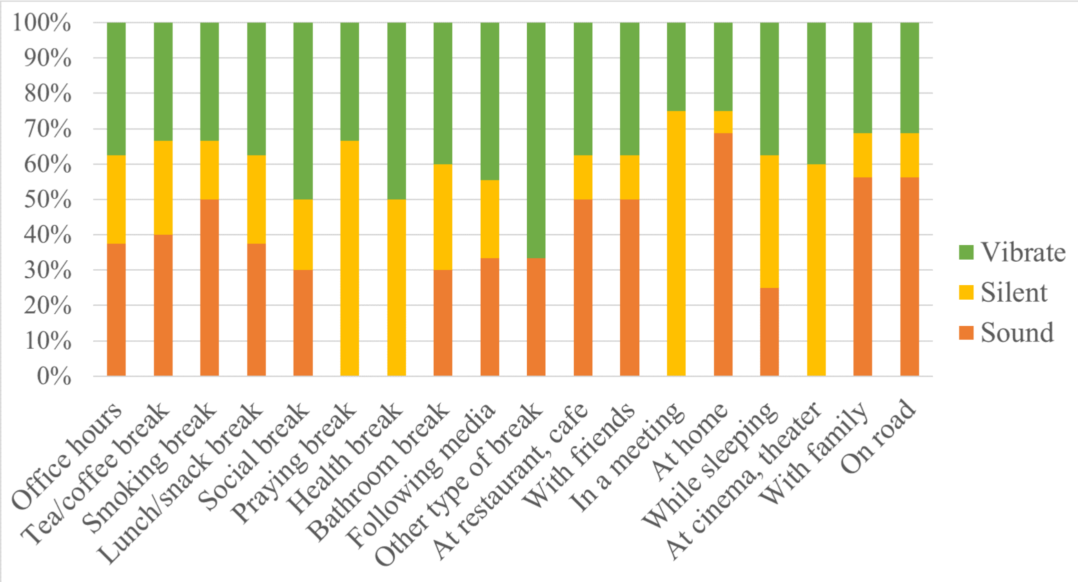


Figure 11: Ringer modes kept in office breaks and different locations

data are combined and named as the *silent-or-vibrate* mode to be used in the modelling phase, and normal mode is modelled as the sound mode.

The ringer mode, in which users their devices during office hours is investigated. The total duration for each ringer mode over the experiment is calculated. The results show that 14 users keep their mobile devices in the sound mode more than 50% of the total duration and the remaining five of them keep in vibrate mode. This finding was also in line with the result obtained from the pre-experiment questionnaire in Figure 11.

As a result, two variables regarding the ringer mode are set: In the first case, "*ringer mode (RM)*" showing whether the phone is in sound mode or silent-or-vibrate mode is used; and in the second case, a variable named "*ringer mode change (RMC)*" is defined. It shows whether the ringer mode has changed compared to the base state, which is determined as the state mostly used during work hours. For this variable, the silent mode is considered as it is (i.e., it has not been merged with the vibrate mode for ringer mode change detection). For example, this variable gets a value of one if the user changed the smartphone's ringer mode setting to a state (e.g., silent mode), which is different from the ringer mode state mostly used (e.g., sound mode). If there is no change compared to the base state, then the variable is set to zero. It is hypothesized that the ringer state change occurs when a significant state change with respect to the user's daily routine is about to happen.

4.5.3 Application Usage (AU)

An application usage session is the time spent between screen-on and off [74, 159–161]. The application usage sessions of each user are extracted in terms of start time, end time, duration of the session (end time - start time) and inter-event times. The sessions are merged where the inter-event time is less than 5 seconds as in previous studies [74, 159]. Thus, it is possible to calculate how long a user used

Table 6: Percentages of number of responses with respect to break availability and current task/break in whole data set

Current Task/Break vs. Break Availability	Cannot take a break	Less than 5-min	5-15 min	More than 15-min
Working	22.35%	13.07%	7.58%	.19%
Tea/coffee break	.19%	3.03%	3.60%	1.14%
Other	2.27%	.95%	.76%	3.03%
Praying break	.19%	.19%	.76%	.76%
Following the media	.00%	.95%	2.09%	.57%
Bathroom break	.00%	1.51%	1.70%	.19%
Health break	.19%	.00%	.19%	.38%
Smoking break	.38%	.57%	.38%	.00%
Social break	.38%	1.33%	3.22%	1.52%
In a meeting	6.25%	.38%	.38%	.00%
Lunch/snack	.57%	1.70%	1.89%	10.41%
On road	.95%	.19%	.95%	.76%

mobile phone before ESM message arrived. The usage sessions with a duration of 5, 10, 15, 30, 45, and 60 minutes before message delivery are investigated.

4.5.4 Break Types vs. Break Availability

Break availability of users is selected as the target variable for prediction. Recall that the participants also stated the ongoing task when notifications arrived. The percentages of the number of responses with respect to break availability and break type categories are given in Table 6. The table shows the relation between break availability and break types. The participants stated that they cannot take a break or they can take up to 5-minutes break when they are in the middle of working (22.35% and 13.07% respectively). Similarly, when they are in a meeting, they stated that they are not available (6.25%). Note that when participants are in a lunch/snack break, they mostly marked the availability of "more than 15 minutes" option (10.41%). It is also seen that the duration of social, bathroom or tea/coffee breaks lasts approximately 5-15 minutes. A chi-square test also showed that break availability and break types are significantly related ($\chi^2(33) = 368.52, p < .001$). Since break availability gives an idea about the break type, the main focus is predicting break availability.

4.6 Hybrid Model

The proposed model is a two-staged model: In the first stage, time is taken into account, then break availability is modelled with time and other variables stated in the previous section. The flowchart of the model is given in Figure 12. The details of the model are given in the subsections.

Before continuing with modelling phase, the data set is split into training and test sets. Repeated random sub-sampling validation (i.e. repeated hold-out) [162] is used. The training and test sets are

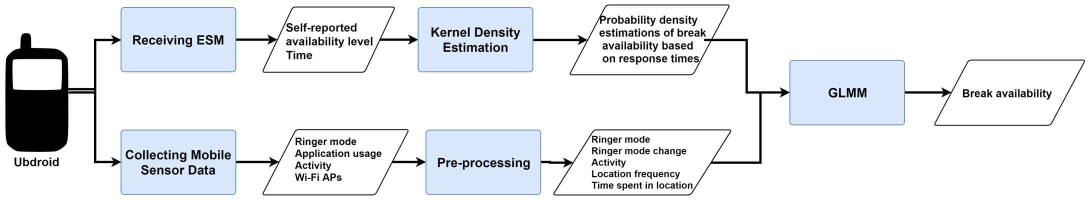


Figure 12: Flowchart of the hybrid model proposed

constructed with the proportions of 30-70, 40-60, 50-50, 60-40, and 70-30. Both sets are randomly formed 20 times. When dividing data into training and test sets, stratified sampling is used so that it enables to balance class proportions in each set.

4.6.1 Modelling Response Time

First, user response time is converted into a numeric variable showing the number of hours in a day. For example; if a user responded to a notification at 12:30:46, the response time in hours is equal to 12.513. Then, the self-reported break availability information of individuals obtained from ESM is considered. This is due to the fact that there are certain time intervals that users prefer to take micro or longer breaks. Kernel density estimation (KDE) is used to estimate the probability density of the break duration versus time. As an example, Figure 13 is given to show a user's KDE plot. The figure shows that the user gives mostly longer breaks (more than 15 minutes) at 12:00. Similarly, the tendency to give shorter breaks (5 minutes or less) is higher between 14:00-16:00. The user does not give breaks before 10:00 or after 17:00.

The *kde* function of the *ks* package in R Software is used with the Gaussian kernel, where the bandwidth is selected with plug-in bandwidth selector. The plug-in bandwidth selector is a highly reliable method for bandwidth selection [163]. Since break availability times and durations vary with each individual, a 2-D KDE is fitted on each user's data set comprising self-reported break availability response with its corresponding time. Then, four new variables are defined regarding break availability predictions called *T1*, *T2*, *T3* and *T4* corresponding to "Cannot take a break", "Less than 5 minutes", "Between 5-15 minutes", and "More than 15 minutes". Finally, for each ESM notification time in the training data set, the probability density function predictions of *T1*, *T2*, *T3* and *T4* were obtained and later included in mixed model analysis.

4.6.2 Modelling Break Availability: GLMM and Comparison

After modelling the response time in the first phase, it is continued with the prediction of break availability with the time variables generated from KDE, and other variables stated in Section 4.5. A Generalized Linear Mixed Model (GLMM) is selected as the main model and the reasons are explained below.

The analyses are performed with the R package named *MCMCglmm* developed by Hadfield [164]. After several trials where it is looked for MCMC convergence and consistency among runs, a burn-in value of 8000, a thinning interval of 50, and the number of MCMC iterations of 50000 were selected.

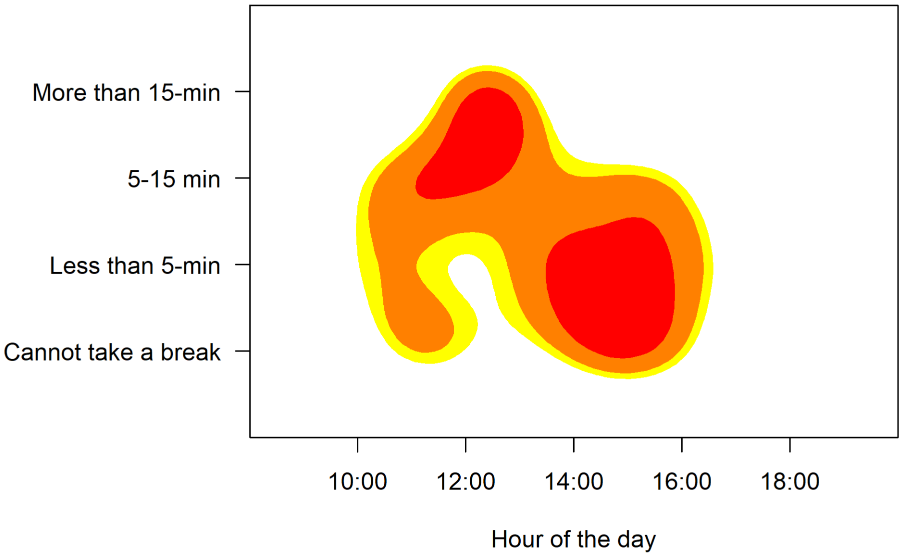


Figure 13: Density plot of a user's break availability levels based on hours of the day

Weakly informative priors for an ordinal response variable are used. The variance component is fixed at one as in [164]. Finally, Gelman-Rubin diagnostics were close to one, and autocorrelation plots were stationary, which means that autocorrelation between consecutive samples in the chain is low enough for the convergence. One of the autocorrelation plots is given in Figure 14 for the variable *location frequency*.

Because the data set consisted of personal data, the variables used in the analyses may be quite different from one individual to another. For example; one may take regular breaks, hence his/her activity switches between "still" to "moving" categories higher than another person who mostly spends his/her time at his/her desk. In such situations, previous studies [18, 34] showed deficiencies on transferring general models to individual-base. Because of that reason, individual models (i.e., trained on only one user's data) can give more accurate results. On the other hand, individual models may suffer from the lack of sufficient individual data for training in the beginning, which is named as the "cold-start problem" [18]. In order to compare the results of GLMM, Random Forest, which is an ensemble learning method [165], is used since this method gives consistently superior results for both individual and general levels in previous studies (e.g. [18, 34]). For training random forest models, both general (with all participants) and individual data (i.e., user-specific models) were used. An individual model is generated for each participant by fitting random forest classifier on each participant's data individually. For all random forest classifiers (general and individual) in the study, the number of trees was selected as 500 since it gave the most accurate results among several values (50, 100, 150, 250, 500, 750, 1000) with cross validation, which was performed on the data set reserved for training. Note that the cross validation is performed on the training data set. The baseline performance is also calculated with the majority classifier that always predicts the class with the highest number of data points.

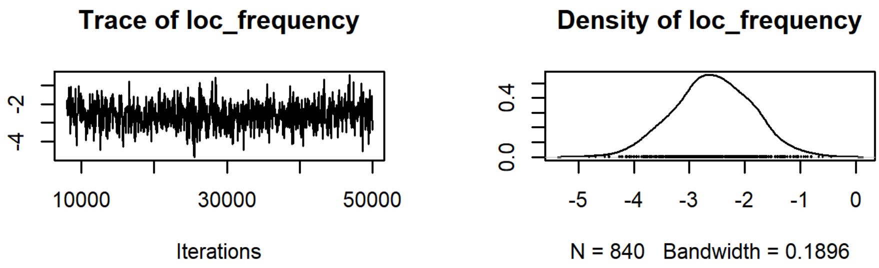


Figure 14: Posterior distribution of a model parameter. Left: Time series of a parameter in the model as MCMC iterates (note that the range of x-values is from 8000 to 50000). Right: probability density estimate of the parameter. The peak of the distribution (the posterior mode) is the most likely value.

4.7 Results

In this section, first, a very brief descriptive statistics regarding the response rates and the availability of user are given. Then, the results of the hybrid model are reported. It is also investigated how the model is comparable to individual and general models, which use the random forest method.

4.7.1 Response Rates and Availability of Users

In total, 921 ESM messages were sent to all users throughout the experiment. The number of 292 messages were sent in the preferred time slots of the users stated in the pre-experiment questionnaire. Ninety-seven of 292 messages were labelled as “can take a break” by the participants whereas 37 of them were labelled as “cannot take a break”, and 158 of the messages were not responded. Similarly, 208 of 629 messages sent in a random time were labelled as “can take a break”, 37 of them were labelled as “cannot take a break”, and 301 of the messages were not responded.

The response rates of each user are given in Figure 15 (top). The response rates are classified as whether the message is sent in preferred time slots or randomly. In the same figure (middle), the positive (i.e., “can take a break”) and negative (i.e., “cannot take a break”) response rates for preferred and random messages are displayed. For example, User 7 responded 11 of the 17 messages sent in preferred times. Six messages were labelled as “can take a break” whereas 5 messages were labelled as “cannot take a break”. The same user responded 17 of 28 messages sent randomly of which 13 messages labelled as “can take a break”, 4 messages as “cannot take a break”. The number of positive responses is higher than negative responses overall. This may be due to the fact that they responded to the messages when they were usually available and ignored the notification messages at other times.

The bottom figure shows each user’s average duration of preferred times per day in minutes. In the pre-experiment questionnaire, some users reported a wider range of time intervals whereas some did not report any such as User 15 and 16. The figures indicate that there is no significant difference between the notifications sent in preferred and random times. This may be attributed to two main reasons. Although some break times are very explicit, such as lunch or praying breaks, other break times may shift or change due to the tasks, which users are carrying out each day. Another reason could be that some users did not think over their daily schedule very well when filling out the questionnaire.

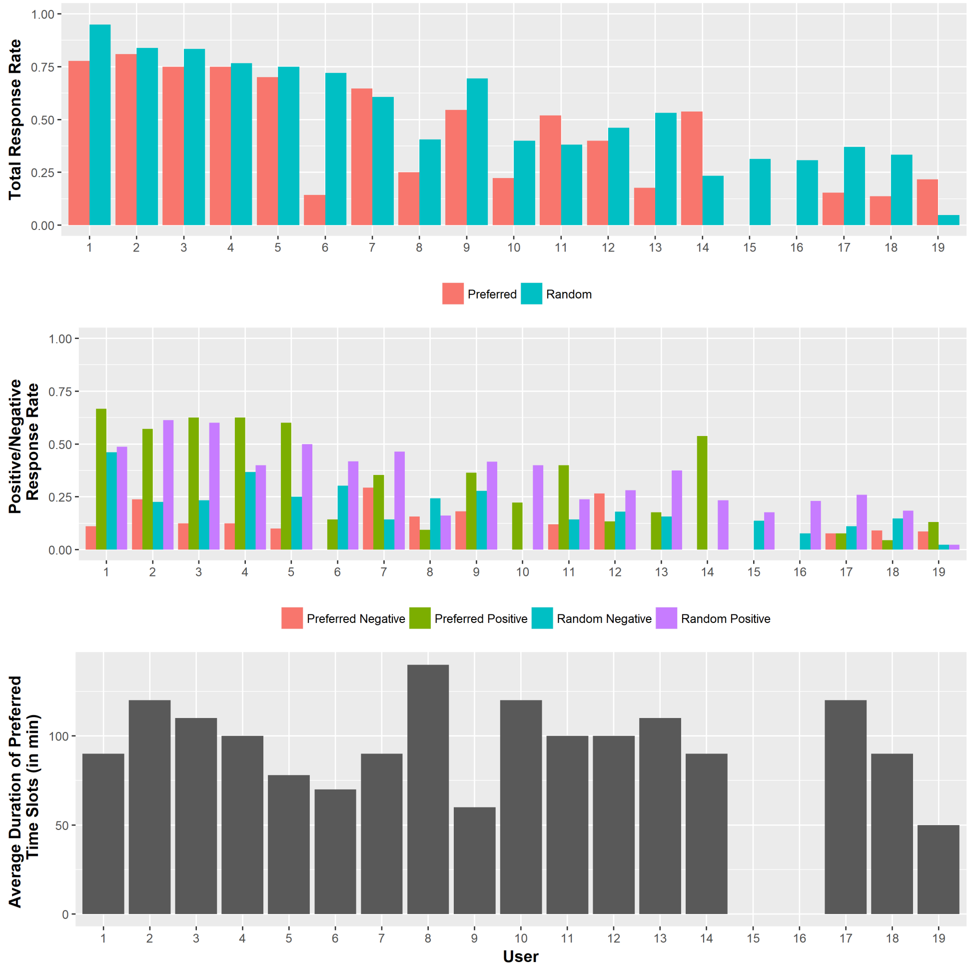


Figure 15: Top: Total response rates of users to the messages sent in preferred and random time slots. Middle: Positive and negative response rates of users to the messages sent in preferred and random time slots. Bottom: Average duration of preferred time slots of users during work hours per day (in minutes).

4.7.2 Model and Feature Set Selection using GLMM

The repeated-measures correlations (rmcorr) are first computed between the predictor variables and break availability. Table 7 shows the correlation results. The results provide an insight about which variables are related to break availability. Rmcorr shows the linear association between the variables and GLMM is inherently a linear model, so the correlation results are taken into consideration for feature set selection instead of variable selection with Gini or other metrics.

As can be seen from Table 7, *location similarity (LS)*, *time spent in location (TSL)*, *location frequency (LF)*, and *5-min application usage* (AU_5) are the most related variables to the break availability. Hence,

Table 7: Repeated measures correlation (rmcorr) coefficients for the predictor variables

		LS	*LF*	*TSL*	AU_5	AU_{10}	AU_{15}	AU_{30}	AU_{45}	AU_{60}
LS	r_{rm}	1	.870	.119	-.120	-.129	-.088	.093	.067	.052
	p		<.001	.007	.006	.004	.05	.04	.13	.24
LF	r_{rm}		1	.114	-.105	-.104	-.079	.065	.048	.028
	p			.01	.02	.02	.08	.14	.28	.53
TSL	r_{rm}			1	-.059	-.065	-.079	-.002	.013	.010
	p				.18	.15	.07	.97	.77	.82
AU_5	r_{rm}				1	.849	.739	.086	.042	.043
	p					<.001	<.001	.05	.35	.34
AU_{10}	r_{rm}					1	.893	.107	.070	.065
	p						<.001	.02	.11	.14
AU_{15}	r_{rm}						1	.134	.092	.083
	p							.002	.04	.06
AU_{30}	r_{rm}							1	.930	.854
	p								<.001	<.001
AU_{45}	r_{rm}								1	.957
	p									<.001
Break	r_{rm}	-.140	-.152	-.226	.071	.066	.019	-.022	-.065	-.054
Availability	p	.001	<.001	<.001	.109	.139	.667	.627	.144	.226

LS: Location similarity, *LF*: Location frequency, *TSL*: Time spent in location, AU_x: Application usage in the last *x* minutes before ESM message

a combination of those variables is included in the GLMM analysis in addition to answer time variables. However, in order to prevent multicollinearity issues, the variable pairs, which have a correlation coefficient with higher than .700, are not included together. *Location frequency* and *location similarity* are highly correlated ($r_{rm} = .870, p < .001$), which is a sign of multicollinearity. Hence, *LF* is selected since it has a higher relation to break availability ($r_{rm} = -.152, p < .001$) than *LS* ($r_{rm} = -.140, p < .001$).

MCMC GLMM fits were built iteratively with the KDE predictions of each break availability level (duration) at a given time *T1*, *T2*, *T3* and *T4* together with *TSL*, *LF*, *activity (A)*, *ringer mode (RM)*, *ringer mode change (RMC)*, and *5-min application usage (AU_5)* parameters, which were given to the models as fixed components (*x* in Equation 1). The random component of the models (*bz* part in Equation 1) refers to the users (19 participants). The response variable (*y* in Equation 1) is the break availability with four levels: "cannot take a break", "less than 5 min", "5-15 min" and "more than 15 min". Note that KDE was fit on the training data set and on the testing data set for each break availability level a KDE prediction was obtained and used in the model. β_0, the intercept in Equation 1 corresponds to the intercept in Table 10.

Table 8 summarizes the models fit for predicting break availability with several combinations of the covariates. The simplest model (Model 1) consisted of *T1*, *T2*, *T3*, *T4*, *TSL* and *LF* as the fixed component. Then, in Model 2, AU_5 was added as another fixed component. In Model 3, *A* is added instead of AU_5. In Models 3 and 4, *RM* and *RM* were added as another fixed component separately and all is included in Model 6 .

Table 8: Models fit upon different covariates for predicting break availability

Model No	**Covariates**
1	$T1 + T2 + T3 + T4 + TSL + LF$
2	$T1 + T2 + T3 + T4 + TSL + LF + AU_5$
3	$T1 + T2 + T3 + T4 + TSL + LF + A$
4	$T1 + T2 + T3 + T4 + TSL + LF + A + RM$
5	$T1 + T2 + T3 + T4 + TSL + LF + A + RMC$
6	$T1 + T2 + T3 + T4 + TSL + LF + A + RM + RMC$

All the models were run on five data sets four of which are sub samples of the original data set because the model and feature selection was desired not to be affected by the variations in the number of responses. The aim was to select the most representative model for all users. Hence, in each sub sample, the data points of the participants with the highest and lowest response rates were eliminated incrementally. To be more specific, the first sample is the full original data set consisting of all the responses of users (*N=528 with 19 users*). In the second data set, two users were removed from the first data set. These users were the ones with the highest and lowest number of ESM responses. In the third data set, two more users with the highest and lowest number of ESM responses were excluded from the second data set. In the fourth data set, two more users with the highest and lowest number of ESM responses were excluded from the third data set. Consequently, the second, third and fourth data sets included *N=469 with 17 users*, *N=410 with 15 users* and *N=350 with 13 users* respectively. Finally, the fifth data set consisted of the users who have Google Activity API in their mobile phones, which means that five users whose activities were predicted were excluded from the first data set (*N=369 with 14 users*). In this way, in sub samples, more balanced data sets were obtained in terms of the number of responses, so that, the model and feature selection process was confirmed as not being affected by the users with the high number of responses. At the end of this process, the model, which gave consistently the lowest DIC on all the data sets, was selected.

The same models were fit in order to investigate the effects of threshold selection using different threshold values when users' locations were clustered. As a result, *LF* and *TSL* variables in aforementioned five data sets were partially changed according to seven threshold values (.05, .10, .15, .20, .30, .40, and .50), which resulted in 35 different runs for each model. For 6 models, it resulted in a total of *35*x*6* runs. Table 9 summarizes these runs with the mean and standard deviation of the DIC values for each data set with different thresholds.

Table 9: Deviance Information Criteria (DIC) estimates for the generalized linear mixed models used to predict the break availability

Model No	**Data Set 1**	**Data Set 2**	**Data Set 3**	**Data Set 4**	**Data Set 5**
1	1239.55 ± 4.70	1101.45 ± 3.66	956.63 ± 3.64	816.77 ± 4.83	834.55 ± 5.22
2	1238.45 ± 2.67	1103.59 ± 4.14	958.71 ± 4.10	816.20 ± 5.05	835.82 ± 2.17
3	1230.29 ± 5.01	1095.76 ± 3.04	951.36 ± 4.03	810.29 ± 5.97	833.89 ± 2.92
4	1225.94 ± 5.47	1088.61 ± 3.49	943.06 ± 2.02	**805.22 ± 1.52**	832.32 ± 1.40
5	1226.35 ± 3.69	1090.11 ± 3.07	947.63 ± 3.61	807.60 ± 3.66	826.96 ± 5.75
6	**1222.03 ± 10.16**	**1088.05 ± 2.73**	**942.26 ± 3.21**	806.66 ± 1.68	**825.93 ± 8.27**

In order to test whether there is a statistical difference between the models, the Friedman Test was applied on the DIC values of the models as suggested in [166]. The Friedman Test [167] is a non-parametric test, and it can be used as an alternative to the repeated-measures ANOVA. The test separately ranks the classifiers for each data set, the classifier, which performs the best gets the first rank, then the second best- performing classifier gets the second rank, and so on [166]. Then, the test determines whether the ranks are statistically different. Similarly, the Wilcoxon signed-rank test [168] is the non-parametric counterpart of the paired t-test. The test also ranks the two classifiers based on their performances, then compares the ranks [166].

In this study, the Friedman Test was used since the normality of the DIC values from four classifiers could not be met. The results show that six models are significantly different ($\chi^2(5) = 116.910, p < .001$). The mean rank of Model 6 is the lowest among all models. Binary comparisons of the models using Wilcoxon signed-rank tests revealed that Model 1 and 2 are not significantly different ($Z = -.966, p = .334$), so Model 1 was selected to continue with since it has the minimum mean DIC between two models. Model 1 and 3 are statistically different ($Z = -4.652, p < .001$), which means that adding *A* to the model explains the variability in the data better. Model 3 and 4 are also statistically different ($Z = -4.815, p < .001$), which means that *RM* is also effective on explaining the variability. Model 4 and 5 are not statistically significant ($Z = -1.294, p = .196$), however, Model 6 is significantly different than Model 4 ($Z = -2.031, p < .05$) and Model 5 ($Z = -2.326, p < .05$). Hence, Model 6 was selected to continue with since it has the minimum mean DIC among all six models.

Table 10 shows the posterior distributions of each parameter in Model 6 with their posterior means, 95% credible intervals (2.5 and 97.5 percentiles of the posterior distribution), and the significance values (*p*). *T1* and *T4* have a significant effect on break availability prediction. It means that KDE break availability predictions for "Cannot take a break" and "More than 15 minutes" obtained using user responses and response time are more effective in predicting the output variable. The inverse relationship between *T1* and the output shows that an increase in the likelihood for the estimation of break availability (*BA*) level 1 (i.e., "Cannot take a break") is a sign of a decrease in the break availability. The positive relationship between *T4* and break availability similarly shows that an increase in the likelihood for the estimation of *BA* level 4 (i.e., "More than 15 minutes") is sign of an increase in the break availability. *T2* and *T3* variables, which are corresponding to *BA* level 2 and 3 respectively, are not significant for predicting break availability levels.

Model outputs show that there is a negative relation between break availability and *LF* (*posterior mean=-.54; 95% CI (-1.11, .04)*, and a significant negative relation between break availability and *TSL* (*posterior mean=-.11; 95% CI (-.22, -.01)*). The results imply that as the location frequency increases, the duration of the breaks decreases or vice versa. Similarly, as the time spent in a location increases, the duration of the breaks decreases or vice versa. Based on the magnitudes, it can be said that *LF* has a higher effect than *TSL*.

In addition to location parameters, the impact of the activity is considered on availability. The model considers "*still*" category as the basis and calculates the posterior mean of the "*moving*" category as .22 with *95% CI= (.09, .37)*. It means that users tend to take longer breaks when they are moving. Similarly, there is an inverse relationship with ringer mode change and break availability. The relationship with ringer mode is also found as positively correlated. Users have a longer break when their phones are in sound mode. The results obtained from the pre-experiment questionnaire support these findings since users mostly change their ringer modes when they attend a meeting or when they do

Table 10: Posterior means, 95% credible intervals (CI) and *p* values of parameters for Model 6

Parameters	Posterior mean	95% CI	p
(Intercept)	2.01	(1.17, 2.92)	<.001
T1	-1.48	(-1.93, -.97)	<.001
T2	-.19	(-.61, .26)	.41
T3	-.23	(-.72, .30)	.34
T4	1.14	(.64, 1.58)	<.001
A[MOVING]	.22	(.09, .37)	.002
RM[SILENT]	-.15	(-.32, .00)	.07
RMC[CHANGE]	-.19	(-.40, .05)	.10
LF	-.54	(-1.11, .04)	.07
TSL	-.11	(-.22, -.01)	.03

not want to be notified in other words the situations when they cannot take a break. In Figure 16, the ringer mode changes in each category of break availability are depicted. The ringer mode at the left hand side of the arrow shows the base ringer mode (ringer mode which user keeps his/her mobile phone in general), whereas the ringer mode on the right shows the current ringer mode when ESM questionnaire is answered. The ringer modes without arrows show the unchanged ringer modes. The figure shows that the ringer state change (specifically vibrate→silent, sound→silent, sound→vibrate) occurs mostly when users cannot take a break and can take less than 5-minute break.

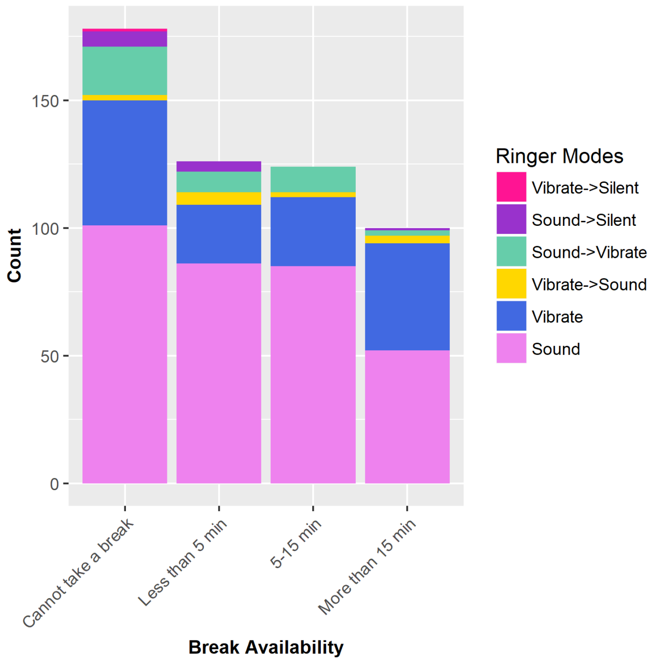


Figure 16: Number of counts of each break availability category grouped by ringer mode change

4.7.3 Comparison Results

As remembered, random forest models both in general (population) level and individual level were fitted for the comparison of GLMM. The variables used in the models were the same with the ones reported in GLMM results: time variables (*T1, T2, T3, T4*), activity, ringer mode, ringer mode change, location frequency and time spent in location. Random forest models were built for each participant (in total 19 models).

The accuracy of the models are reported as the performance metric. The estimated accuracy is obtained by averaging 20 runs. The only difference for the individual random forest classifier is that the data set consisted of one user's responses. In Table 11, the mean and standard deviations of accuracy values obtained from four different classifiers for predicting break availability are presented. The percentage of the training set is reported in the first column.

Table 11: Comparison of model accuracy values for predicting break availability

Training Percentage	**GLMM**	**General Random Forest**	**Individual Random Forest**	**Baseline**
30%	53.47% ± 12.70%	45.73% ± 14.47%	33.42% ± 14.10%	33.87% ± 12.06%
40%	52.86% ± 13.36%	46.69% ± 14.50%	33.16% ± 14.61%	34.28% ± 11.89%
50%	52.36% ± 14.13%	44.95% ± 15.76%	34.38% ± 15.03%	33.15% ± 13.11%
60%	53.07% ± 16.19%	46.99% ± 17.99%	33.96% ± 18.00%	33.86% ± 15.09%
70%	53.51% ± 19.88%	46.79% ± 19.74%	34.42% ± 19.07%	31.51% ± 16.46%

As can be seen from the table, GLMM predicted break availability levels better than the general random forest model, individual random forest and baseline classifiers. Since the Shapiro-Wilk Test showed that the accuracy values of the models did not distribute normally ($p < .001$), the Friedman Test was conducted on the accuracy values obtained from all runs (5 different training percentages x 20 runs x 19 users = 1900 accuracy values for each classifier). The results of Friedman Test showed that the average accuracy values obtained from four classifiers are significantly different for the prediction of break availability ($\chi^2(3) = 1908.315, p < .001$). Then, Wilcoxon signed-rank tests were conducted for binary comparisons of the models as post-hoc tests. The accuracy obtained from GLMM is significantly higher than the general random forest's accuracy ($Z = -18.667, p < .001$), individual random forest's accuracy ($Z = -30.004, p < .001$), and the baseline accuracy ($Z = -31.238, p < .001$).

4.7.4 Individual Models

The accuracy values of each user's model are reported with their means and standard deviations in Table 12. Seventy percent training data was used for the runs reported in the table, and in total, 20 runs were made. The bold values in the table show the highest accuracy among the four classifiers. The results are given based on the participants' number of responses (*N*) in descending order. The participants whose number of responses is less than 20 are not included in the table because such a limited number of data points might not be sufficient for individual models to learn the target category.

The accuracy values obtained from individual random forest classifier were not as high as the ones obtained from GLMM and general random forest classifier most of the time. Individual random forest

Table 12: The average accuracy and standard deviation values for predicting break availability levels with GLMM, general random forest, individual random forest classifier, and baseline classifier

User No	Number of Responses	GLMM	General Random Forest	Individual Random Forest	Baseline
U15	50	**51.25% ± 11.74%**	31.25% ± 9.78%	13.75% ± 10.28%	26.25% ± 8.54%
U17	49	**53.21% ± 9.33%**	40.36% ± 12.70%	33.57% ± 9.14%	37.14% ± 9.25%
U16	47	**50.33% ± 13.80%**	43.00% ± 13.17%	45.33% ± 10.13%	25.33% ± 11.49%
U03	45	61.70% ± 11.42%	**61.79% ± 7.06%**	61.07% ± 9.12%	43.57% ± 10.59%
U06	37	**50.00% ± 14.89%**	46.82% ± 12.26%	43.64% ± 13.06%	27.73% ± 13.02%
U12	36	**50.91% ± 11.94%**	39.09% ± 11.83%	34.09% ± 9.27%	49.09% ± 14.86%
U07	32	53.50% ± 15.65%	**56.50% ± 15.31%**	52.00% ± 15.08%	26.50% ± 8.13%
U01	31	40.56% ± 13.62%	**46.67% ± 13.77%**	38.33% ± 13.23%	19.44% ± 7.10%
U04	31	**47.78% ± 12.54%**	37.78% ± 13.20%	23.89% ± 9.72%	31.11% ± 10.57%
U13	27	**51.88% ± 14.21%**	40.63% ± 17.62%	31.25% ± 13.75%	45.00% ± 14.28%
U08	25	**50.71% ± 15.00%**	47.86% ± 18.11%	30.71% ± 16.24%	40.71% ± 21.88%
U05	24	**55.00% ± 18.11%**	32.14% ± 15.28%	35.71% ± 15.72%	17.14% ± 8.79%

classifier made worse predictions even than the baseline classifier for six users. Furthermore, GLMM predicted nine users' break availability the most accurately among four classifiers. GLMM appears to be a more appropriate method to predict break availability of the participants.

4.8 Discussion

In this section, the findings are discussed with respect to each feature of the model in detail. The results of the comparison are elaborated.

4.8.1 Time

Previous studies included time variable as a hour of day solely in their models [34, 139, 140], or or as a part of a day (e.g. morning, evening) [143, 144, 169]. As a difference from previous work, time was included as a density of each break availability level. A 2-D kernel density estimation was employed upon hour of day and break availability level. This conversion also facilitated the modelling of time using generalized linear mixed models. As a consequence, a hybrid model, which has not been employed in the previous studies, was built. Based on the results, time appears to be an effective factor on predicting break availability. Specifically, knowledge workers tend to give longer breaks especially at midday (between 12-2 PM). They also tend to give shorter breaks at about 2-4 PM.

4.8.2 Location

Previous studies showed that location is an important context information for interruptibility [7,72,118, 137, 139, 140, 144, 170]. The effects of the location were investigated in different measures: location frequency (i.e. how frequent a user visits a place), time spent in a location (i.e., how many minutes the user spent in a location) and location similarity (how similar a place is to the user's workplace). Based on the results, location frequency and time spent in location explained break availability of knowledge workers.

The results show that users tend to have longer breaks at locations where they visit less frequently. These locations maybe the places where they have lunch breaks. The preliminary results showed that participants have longer breaks when they are in lunch/snack breaks. Hence, the model output supports this finding. Similarly, when they are at the beginning of arrival to a location, they tend to give longer breaks. To our knowledge, this study is the first, which uses location parameters in terms of duration, frequency and similarity metrics for predicting the availability of knowledge workers.

4.8.3 Physical Activity

The activity of users (whether the user is *still* or *moving*) affects the break availability as found in the previous studies [1, 143, 144, 171, 172]. The results are in line with these findings. It is found that users are more likely to have a break when they are already moving. When they are not moving, the tendency of not taking a break is higher.

4.8.4 Ringer Mode

In the pre-experiment questionnaire, users stated that they keep their mobile phones in different ringer modes in different contexts as in [72]. Original ringer mode values are included in the analyses, but different from previous work, a variable was kept for ringer mode change and investigated its effect. According to the model, including ringer mode and/or ringer mode change in the models is essential for explaining break availability. Users tend to not take a break when the ringer mode is in silent-or-vibrate mode, and when there is a change in ringer mode. Since most of the users keep their mobile phones in sound mode, the change may mean that they do not want to be interrupted. Hence, ringer mode change to vibrate or silent is an indicator of unavailability.

4.8.5 Application Usage

Previous work showed that an increase in application usage may be an indicator for an opportune moment to take a break or responsiveness [74, 173]. The use of 5,10, 15, 30, 45 and 60 minutes before notification arrival was extracted and given to the models. According to the results, the models based on application usage had lower DIC values, which means that application usage is one of the explanatory variables for the break availability. The application usage variables did not have a significant effect on break availability ($p > .05$) in the models, which include those variables.

Previous work showed the importance of application usage on the interruptibility, however, in this study, it did not have a significant effect. This may be due to cultural or environmental factors. The data was collected in office environments during work hours, hence employees might not be able to use their mobile phones even when they are available for a break. In addition, other factors that have not been included in the study such as application type could be a reason for the result. Notification may have arrived at the moment when users take a note on their mobile phones during a meeting. In such situations, application type or category may become more important than duration of application usage. Sahami Shirazi et al. [142] had found similar results showing the importance of application type on susceptibility.

4.8.6 Individual Models and Model Comparison

GLMM was compared with random forest classifier, which is commonly used in interruptibility domain (e.g. [18, 34]), and also the baseline classifier. The results show that GLMM may be preferred compared to individual random forest models when sufficient data is not available. GLMM handles the data insufficiency, and it incorporates random effects, which means that it fits both an individual and a general-level mean. Hence, when there is a totally new user with an insufficient number of data points, the general-level mean might be used for that user at first until his/her data is high enough to fit an individual-level mean. That brings a solution to the cold-start problem stated in [18]. In their study, individual random forest model gives as accurate result as general random forest model when the training set has 45 samples. Their individual model gives more accurate results after 45 data points in the training set. The maximum number of points per user in this study is around 50. The reason of under-performance of individual models compared to GLMM or general models could be the limited number of points. As a consequence, GLMM might be considered as a good solution specifically when the data points are below 40.

4.8.7 Break Types vs. Break Availability

The data showed that certain types of breaks have different characteristics in terms of duration. Lunch/snack breaks last more than 15 minutes, whereas social or coffee/tea breaks last approximately 10-15 minutes. Users marked themselves “not available” when they are in the middle of working or meeting. Therefore, it is possible to infer what users may be busy with if break availability can be predicted. However, the use of break types for prediction was not successful with GLMM due to the limited samples and similar characteristics of certain break types. Therefore, break availability was studied in this work. The results of the study are inline with a very recent study of receptivity for health interventions [118]. As differently, only the work-related rest breaks were considered in this study, whereas they included a variety of activities such as resting, relaxing, or watching video games.

CHAPTER 5

SOCIAL-CONTEXT AND PERSONAL NORMS IN OFFICE SETTINGS

In this chapter, the effects of social context, personal norms and mobile phone usage related factors on the inference of work engagement/challenge levels of knowledge workers and their responsiveness to health-related notifications are investigated. The following sections explain the method used in the second study of the thesis. The results are presented and discussed in the last sections.

5.1 Introduction

In an office setting, there are many factors that can affect the responsiveness of office workers to health intervention messages, in particular their *engagement* and *challenge* levels [2, 19]. The responsiveness can also be affected by the health history of users. In the mobile context, users may be more likely to respond to the messages sent by break reminder applications, if they experienced musculoskeletal discomfort due to their sedentary life. Higher awareness or self-regulation may increase the responsiveness to break-reminder notifications [26]. Besides, social factors, such as subjective norms, have been found as a precursor related to behavioral intention [27, 28]. Recent studies showed that office employees are influenced by their co-workers regarding prolonged sitting behavior [30] or performing physical activity [31]. Hence, office workers might also be influenced to take rest breaks by their colleagues, and that affects the responsiveness to the mobile rest breaks reminders. Finally, the number of colleagues in the same office might be another factor for both responsiveness to the reminders. It has been showed that office type (shared or private offices) has a significant effect on distractions [32] and also on sitting time [33].

In the study in this chapter, a framework related to the responsiveness of office workers is proposed. The design, implementation and evaluation of the framework for the inference of engagement/challenge levels of office workers and their responsiveness to well-being related mobile notifications are given as a contribution to the studies above. The responsiveness is investigated with several metrics such as acquiescence, disacquiescence, and extreme response style (negative and positive). Mainly, the main target of the study is offering a solution for the second research question: *"How are the musculoskeletal discomfort, awareness, office-related factors, personality traits, and mobile application usage of office workers related to their responsiveness to break-reminder notifications?"*.

5.2 Method

Based on the studies given above, a research framework is proposed for understanding the responsiveness with the variety caused by different individuals, and it is given in Figure 17. The responsiveness metrics were hypothesized to be affected by awareness of rest breaks, musculoskeletal discomfort, personality traits, mobile application usage, and office-related factors. The details of the measures are given in following sections. The main steps of conducting this study are given in Figure 18.

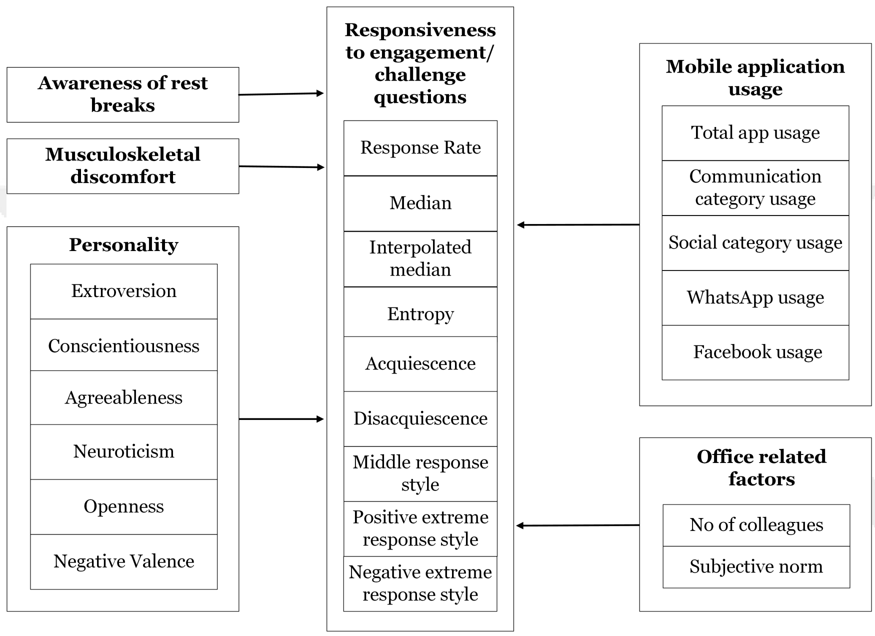


Figure 17: Research framework proposed in the second study

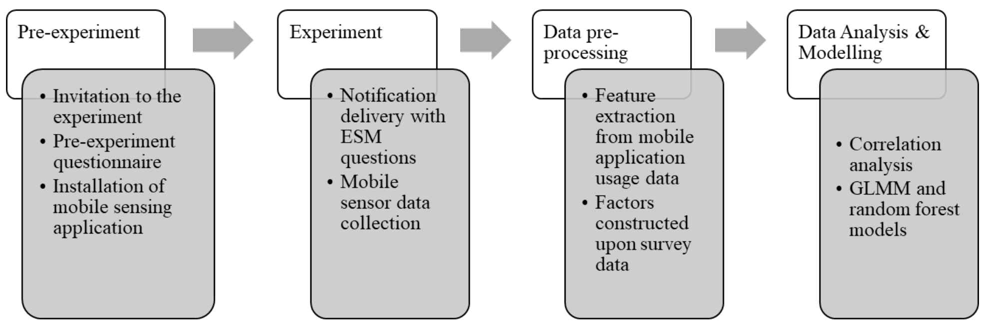


Figure 18: Main steps of the second and the third studies

5.3 Data Collected and Feature Extraction

In this section, the descriptive statistics of the data collected are presented. Besides, the features used in the analyses are described.

5.3.1 Application Usage

An application usage session is defined as the time spent between the screen on and off [74,159]. Based on this definition, the application usage sessions of each participant were extracted. Each session's start-end time, duration and the inter-event time information were recorded. Inter-event time refers to the duration between two consecutive usage sessions, i.e., the time difference between the previous session's end time and the current session's start time. The sessions where inter-event time is less than 5 seconds were merged as suggested in the previous studies [74, 159]. The usage sessions with a duration of 5, 10, 15, 30, 45 and 60 minutes before each message delivery were investigated from the core participants from whom application package names were obtained.

In Table 13, the descriptive statistics (mean, median and standard deviation) for the application usage in the last 5, 10, 15, 30, 45 and 60 minutes before ESM messages were responded, is shown. The number of sessions recorded from 19 participants (those who had a response rate of 25% or higher in ESM questions) is 528 in total. In the 5-min case, more than half of the data set consisted of zeros, which means that more than half of the ESM messages were responded when users did not use their mobile phone in the last 5 minutes. As the duration increases, the number of sessions with no application usage decreases (39.47% for the 10-minutes, 29.43% for the 15-minutes, 20.10% for the 30-minutes, 12.68% for the 45-minutes, and 9.22% for the 60-minutes).

Table 13: The descriptive statistics of the application usage (in seconds) in the last 5, 10, 15, 30, 45 and 60-minutes before ESM messages were responded

Time Window	Mean	Std. Dev	Median	Number of sessions with no application usage	Percentage of sessions with no application usage
5 minutes	20.91	35.42	2.00	250	47.34%
10 minutes	31.37	49.93	11.00	200	37.88%
15 minutes	43.04	61.70	21.50	157	29.73%
30 minutes	395.40	508.58	169.50	29	5.49%
45 minutes	596.49	730.86	297.00	22	4.17%
60 minutes	791.29	934.63	448.50	13	2.46%

The usage of applications per category was also considered. However, only from 14 number of participants such data could be obtained. The number of sessions recorded from 14 participants (those whose application package names could be obtained and had a 25% or higher response rate in ESM questions) is 418 in total. The application categories were matched using the Google Play Store. The category usage in the last 5, 10, 15, 30, 45 and 60 minutes before each ESM message responded was calculated. The descriptive statistics (mean, median and standard deviation) of each category usage in the last 60 minutes is given in Table 14. The number of zeros (i.e. non-used categories) and their percentages were also presented. As can be seen from the table, the category usage data is sparse. The densest category is communication; 46.17% of the sessions have communication applications usage greater than zero. The second densest category is social category, with a percentage of 59.33. The remaining categories have less than 30% fullness (i.e. their usage is equal to zero for 70% or higher). For this reason, communication and social categories were used in the analyses.

Table 14: The descriptive statistics of the application category usage (in seconds) in the last 60-min before ESM messages were responded

Category Name	Mean usage	Standard Deviation	Median	Number of sessions with no application usage	Percentage of sessions with no application usage
Communication	186.57	273.21	78.5	67	46.17%
Social	69.08	158.44	0	248	59.33%
Tools	14.04	93.77	0	324	77.51%
Productivity	3.28	16.43	0	375	89.71%
Finance	8.54	45.89	0	388	92.82%
Photography	7.11	47.56	0	388	92.82%
Personalization	4.14	28.06	0	391	93.54%
Lifestyle	82.46	407.37	0	392	93.78%
Game	23.43	121.45	0	393	94.02%
News and Magazines	13.15	101.47	0	395	94.50%
Food and Drink	3.53	28.40	0	400	95.69%
Music and Audio	2.57	18.16	0	400	95.69%
Travel and Local	4.27	58.50	0	401	95.93%
Video Players and Editors	6.04	70.71	0	403	96.41%
Business	1.56	14.14	0	408	97.61%
Weather	.41	3.81	0	409	97.85%
Shopping	.88	11.11	0	410	98.09%
Entertainment	7.63	73.21	0	411	98.33%
Books and Reference	1.58	31.08	0	414	99.04%
Sports	1.04	18.13	0	414	99.04%

The cumulative application usage in hours per user during work hours was calculated. A total of 24 users among all 31 participants were considered since only these participants gave access to the application to collect their mobile application usage details in the background. Note that only 14 out of them had a response rate of 25% or higher in ESM questions. The descriptive statistics of the aggregated mobile application usage parameters are given in Table 15.

- *Total Application Usage (TAU)*: shows the total duration of application usage.
- *Total Communication Category Usage (TCOM)*: denotes the total duration of application usage in the communication category.
- *Total Social Category Usage (TSOC)*: denotes the total duration of application usage in the social category.
- *Total Facebook Usage (TFB)*: shows the total duration of Facebook usage.
- *Total WhatsApp Usage (TWA)*: shows the total duration of WhatsApp usage.

Table 15: The descriptive statistics of the constructs with continuous parameters used in the analyses

Construct Name	Parameter Name	N	Min.	Max.	Mean±SD	Median
Responsiveness	Response Rate	31	.09	.94	.42±.27	.35
	Engagement					
	Median	31	1.00	5.00	2.82±1.01	3.00
	Interpolated Median	31	1.23	4.63	2.80±.89	2.81
	Entropy	31	.72	2.29	1.79±.37	1.86
	Polarization	31	.05	.79	.39±.17	.40
	Acquiescence	31	.00	.80	.36±.22	.37
	Disacquiescence	31	.00	.82	.45±.22	.46
	Acquiescence Balance	31	-.80	.80	.09±.41	.14
	Middle Response Style	31	.00	.57	.18±.14	.18
	Positive Extreme Response Style	31	.00	.57	.19±.17	.15
	Negative Extreme Response Style	31	.00	.68	.27±.16	.28
	Challenge					
	Median	31	1.00	4.00	2.29±.79	2.00
	Interpolated Median	31	1.18	3.60	2.30±.67	2.38
	Entropy	31	.83	2.23	1.68±.38	1.69
	Polarization	31	.07	.69	.29±.15	.29
	Acquiescence	31	.00	.56	.20±.16	.19
	Disacquiescence	31	.00	1.00	.58±.25	.52
	Acquiescence Balance	31	-.56	1.00	.38±.39	.38
	Middle Response Style	31	.00	.63	.23±.17	.14
	Positive Extreme Response Style	31	.00	.36	.06±.09	.00
	Negative Extreme Response Style	31	.00	.74	.33±.17	.33
Mobile application usage (in minutes)	Total application usage	31	1.55	65.14	15.16±12.88	13.44
	Total social category usage	24	.01	4.07	1.69±1.35	2.09
	Total communication category usage	24	.36	20.33	7.24±5.34	6.29
	Total Facebook usage	24	.00	4.20	.90±1.07	.75
	Total WhatsApp usage	24	.00	6.00	1.74±1.43	1.59
Personality	Extroversion	19	2.13	4.63	3.18±.67	3.13
	Conscientiousness	19	1.38	4.38	3.45±.73	3.63
	Agreeableness	19	3.63	5.00	4.24±.36	4.25
	Neuroticism	19	1.89	3.78	2.80±.52	3.00
	Openness	19	2.67	4.50	3.44±.54	2.63
	Negative Valence	19	1.00	2.67	1.64±.56	1.13
Office related factors	Number of colleagues in office	31	1.00	50.00	10.13±14.51	3.00
	Subjective Norm	31	2.00	9.00	6.45±1.97	7.00

5.3.2 Responsiveness Metrics

The responsiveness of participants were explored in detail. The parameters explained below were calculated based on engagement and challenge responses from ESM questions and aggregated. The descriptive statistics of the parameters are presented in Table 15.

- *Response Rate*: is calculated by dividing the number of each participant's responses to the total number of ESM messages sent to that participant.
- *Median of Engagement and Median of Challenge*: refer to the median value of the responses of each participant on engagement and challenge related ESM questions respectively.
- *Interpolated Median of Engagement and Challenge*: Medians may suffer from ignoring the weights caused by the distributions of responses above or below the median. The interpolated medians take into account the number of data points, which are strictly below or above the median. Hence, in this study, the interpolated medians of engagement and challenge responses were calculated respectively.
- *Entropy of Engagement and Entropy of Challenge*: refer to the Shannon entropy of the responses of each participant on engagement and challenge related ESM questions respectively. The formula of the entropy is given in Equation 6. In the formula p_i refers to the proportion of item *i* (where *i=1,..,5* since engagement and challenge levels were measured with 5-level Likert scale). A lower entropy indicates higher homogeneity of responses, whereas a higher entropy indicates higher heterogeneity of responses. For example; if a participant's responses are in only one category (e.g., the user selected 3 for all engagement questions), then his/her entropy of engagement will be zero which indicates pure homogeneity. On the other hand, entropy gets higher values as the responses fall into different categories.

$$Entropy = \sum_{i=1}^{5} -p_i log_2 p_i \tag{6}$$

- *Polarization of Engagement and Polarization of Challenge*: refer to the polarization of the responses of each participant on engagement and challenge related ESM questions respectively. Polarization is defined as $(1 - agreement)/2$. Details of the agreement calculation is given in [174]. A polarization score of zero indicates the responses were gathered in one category whereas a score of .50 implies the responses are almost uniformly scattered among the categories. A polarization score of one shows the responses are divided in few non-neighbouring categories.
- *Acquiescence of Engagement and Challenge*: refer to the tendency to be highly engaged or challenged with work (i.e. giving 4 or 5 to the engagement/challenge level responses). It is calculated by dividing the number of responses recorded as 4 or 5 by the total number of responses.
- *Disacquiescence of Engagement and Challenge*: refer to the opposite of the acquiescence, which means the tendency to be low engaged or challenged with work (i.e. giving 1 or 2 to the engagement/challenge level responses). It is calculated by dividing the number of responses recorded as 1 or 2 by the total number of responses.
- *Acquiescence Balance of Engagement and Challenge*: imply the difference between acquiescence and disacquiescence.
- *Middle Response Style of Engagement and Challenge* : refer to the proportion of the responses that received middle (3) response.
- *Positive Extreme Response Style of Engagement and Challenge*: Positive extreme responses imply the responses with the category of 5. Hence, the positive extreme response style indicates the proportion of the responses that received positive extreme responses among all responses.

Table 16: The descriptive statistics of the indicators with ordinal parameters used in the analyses

Construct Name	Parameter Name	Values	Frequency	Percentage
Awareness about rest breaks	Taking regular rest breaks	1	3	9.7%
		2	18	58.1%
		3	10	32.3%
	Doing office exercises	1	20	64.5%
		2	6	19.4%
		3	5	16.1%
Musculoskeletal discomfort	Feeling pain/numbness	1	8	25.8%
		2	5	16.1%
		3	18	58.1%

- *Negative Extreme Response Style of Engagement and Challenge*: Negative extreme responses imply the responses with the category of 1. Hence, the negative extreme response style indicates the proportion of the responses that received negative extreme responses among all responses.

5.3.3 Survey Data Set

As remembered, personality traits were collected with BPTI, and office-related factors, musculoskeletal discomfort, and awareness about rest breaks were collected with the pre-experiment survey. The features and their descriptive statistics obtained from the survey results are given in Table 15 and in Table 16. The total number of participants in the data set is 31. Nineteen of users filled the BPTI at the end of the experiment. The following features were calculated for each user:

- *Personality*: scores obtained from the BPTI show the degree on each trait. Extroversion, Conscientiousness, Agreeableness, Neuroticism, Openness, and Negative Valence scores.
- *Awareness about rest breaks*: was measured with the degree of participants taking rest breaks and doing office exercises.
- *Musculoskeletal discomfort*: indicates the degree of feeling pain/numbness when working in front of computers in the office.
- *Office-related Factors*: refers to the number of colleagues in each participant's office, and subjective norm, which identifies the degree of participants affected by their colleagues for taking rest breaks.

5.4 Relational Analysis between the Constructs in the Survey Data Set

In this section, the relations among the constructs in the survey data set (personality, musculoskeletal discomfort, awareness about rest breaks, and office-related factors), mobile application usage parameters, and responsiveness metrics were investigated using Kendall's Tau correlation based on users' aggregated values.

Table 17: Kendall's Tau correlation coefficients for the responsiveness variables, musculoskeletal discomfort, awareness about rest breaks, and office-related parameters

		Feeling pain/ numbness	Taking regular rest breaks	Doing office exercises	Subjective norm	No of colleagues
Response	τ	.13	.04	-.12	.02	.08
Rate	p	.36	.79	.41	.90	.52
Median of	τ	-.07	-.09	-.01	.14	-.14
Engagement	p	.67	.58	.93	.34	.34
Interpolated Median	τ	.16	-.20	-.13	-.06	.03
of Engagement	p	.26	.17	.36	.65	.84
Entropy of	τ	.24	-.01	.13	.14	.14
Engagement	p	.09	.95	.39	.32	.28
Polarization of	τ	.16	-.06	.11	.00	.12
Engagement	p	.26	.68	.47	.97	.37
Acquiescence of	τ	.17	-.15	-.15	-.08	-.06
Engagement	p	.25	.31	.29	.56	.63
Disacquiescence	τ	-.16	.11	.17	-.01	-.10
of Engagement	p	.29	.45	.23	.93	.45
Acquiescence Balance	τ	.16	-.15	-.17	-.06	.02
of Engagement	p	.28	.29	.24	.68	.90
Middle RS	τ	-.01	.08	-.05	.16	.16
of Engagement	p	.95	.61	.75	.25	.25
Positive Extreme RS	τ	.06	-.20	.02	-.19	-.12
of Engagement	p	.68	.19	.89	.17	.39
Negative Extreme RS	τ	-.17	.25	-.03	-.01	.05
of Engagement	p	.24	.09	.82	.92	.72
Median of	τ	-.14	.16	-.13	-.03	.09
Challenge	p	.37	.30	.43	.83	.52
Interpolated Median	τ	.00	-.09	-.01	-.04	.17
of Challenge	p	.98	.56	.92	.79	.21
Entropy of	τ	.16	.07	.15	.16	.23
Challenge	p	.26	.61	.30	.25	.09
Polarization of	τ	.14	-.01	.00	-.02	.12
Challenge	p	.32	.97	1.00	.89	.38
Acquiescence of	τ	.13	-.11	-.10	-.08	.05
Challenge	p	.38	.47	.52	.57	.73
Disacquiescence	τ	.03	.04	.02	.10	-.13
of Challenge	p	.82	.80	.89	.47	.33
Acquiescence Balance	τ	.05	-.09	-.04	-.10	.09
of Challenge	p	.76	.55	.79	.47	.50
Middle RS	τ	-.09	.05	.03	.02	.11
of Challenge	p	.55	.75	.86	.90	.41
Positive Extreme RS	τ	-.03	.00	-.09	-.23	-.07
of Challenge	p	.86	1.00	.58	.12	.65
Negative Extreme RS	τ	-.02	.14	-.08	-.01	-.06
of Challenge	p	.89	.33	.60	.96	.67

Table 17 depicts the correlation results between the responsiveness metrics and the factors of musculoskeletal discomfort, awareness about rest breaks, and office environment as a part of the second research question. The musculoskeletal discomfort was assessed as the degree of feeling pain/numbness while working on computers. The results show that the degree of feeling numbness and pain during work is positively related to the entropy of engagement ($\tau = .25, p = .09, N = 31$). The engagement responses of the participants who felt a higher amount of musculoskeletal discomfort were more heterogeneous than the ones who felt less amount of musculoskeletal discomfort. In other words, the participants who suffered more from musculoskeletal discomfort responded with a higher number of categories as their engagement levels.

The awareness was measured with the degree of taking regular rest breaks, and the degree of doing office exercises. Based on the results, taking regular rest breaks is positively related to the negative extreme response style of engagement ($\tau = .25, p = .09, N = 31$). It can be inferred that the participants who give rest breaks more frequently selected the response item "I am not engaged at all" higher number of times than the ones who give rest breaks less frequently. Even though it was not hypothesized, an interesting result has been found: taking rest breaks is negatively related to the feeling pain/numbness ($\tau = -.50, p = .003, N = 31$). It means that the participants who give rest breaks more frequently felt less musculoskeletal discomfort while working than the ones who give rest breaks less frequently.

The office-related factors were measured by the number of colleagues and subjective norm. The results showed that the number of colleagues is in a weak positive relationship with the entropy of challenge ($\tau = .23, p = .09, N = 31$). It means that the participants who share their offices with a higher number of colleagues responded challenge questions more heterogeneously (i.e., responded with a higher number of item categories) than the ones who share their offices with a lower number of colleagues. At the same time, subjective norm is significantly related to the number of colleagues in office ($\tau = .35, p = .01, N = 31$). Again, the participants who share their offices with a higher number of colleagues stated that they are affected more from their colleagues regarding giving rest breaks than the ones who share their offices with a lower number of colleagues.

Table 18 presents the correlation coefficients calculated upon personality traits and responsiveness constructs as another part of the second research question. The results are as follows:

- *Extroversion*: trait is negatively related to the entropy of challenge levels ($\tau = -.31, p = .07, N = 19$). It indicates that extroverts have more homogeneous responses than introverts. Similarly, it has a negative relation with positive extreme response style of challenge ($\tau = -.31, p = .09, N = 19$). It can be inferred that the proportion of responding "I am very challenged" item is higher for introverts than for extroverts.

- *Conscientiousness*: trait has a higher number of significant relations with responsiveness parameters than other traits. It is in a positive relation with interpolated median of engagement ($\tau = .47, p = .007, N = 19$), with interpolated median of challenge ($\tau = .35, p = .04, N = 19$), with acquiescence of engagement ($\tau = .38, p = .03, N = 19$), with positive extreme response style of engagement ($\tau = .40, p = .02, N = 19$), with acquiescence of challenge ($\tau = .33, p = .06, N = 19$), with middle response style of challenge ($\tau = .30, p = .08, N = 19$), and with positive extreme response style of challenge ($\tau = .34, p = .07, N = 19$). As expected, it is in a negative relation with the opposite of those metrics: namely with disacquiescence of engagement ($\tau = -.29, p = .09, N = 19$), with negative extreme response style of engagement

($\tau = -.44, p = .01, N = 19$), with disacquiescence of challenge ($\tau = -.35, p = .04, N = 19$), and with negative extreme response style of challenge ($\tau = -.35, p = .04, N = 19$). Those results infer that conscientious participants tend to give higher scores to engagement and challenge level questions than the participants who are low on conscientiousness.

- *Agreeableness*: trait and the median of engagement/challenge levels are negatively related. The strength of the relation between Agreeableness and challenge levels is higher ($\tau = -.42, p = .03, N = 19$) than the relation between Agreeableness and engagement levels ($\tau = -.33, p = .08, N = 19$). It implies that the more agreeable a person is, the lower the level of engagement and challenge scores they reported.

- *Neuroticism*: was not found related to any of the parameters.

- *Openness*: is weakly negatively related with entropy parameters ($\tau = -.33, p = .06, N = 19$ for engagement, $\tau = -.49, p = .005, N = 19$ for challenge). Like Extroverts, the participants high on Openness trait tend to have more homogeneous responses than people low on Openness trait. In addition, Openness is found to be negatively related to the response rates of the participants ($\tau = -.38, p = .03, N = 19$). The participants high on Openness responded to a lower number of messages than the ones who low on Openness.

- *Negative Valence*: is positively related to the median of challenge ($\tau = .49, p = .009, N = 19$). The relatedness of Negative Valence to entropy parameters is also in a positive way ($\tau = .33, p = .06, N = 19$ for engagement, $\tau = .34, p = .05, N = 19$ for challenge). Similarly, it is in a positive relation with positive extreme response style of challenge ($\tau = .51, p = .006, N = 19$). Those results suggest that participants who perceive themselves more negatively responded engagement and challenge levels higher than the ones who perceive themselves more positively.

The correlation coefficients between application usage parameters of total application usage, total social category usage, total communication category usage, total Facebook usage and total WhatsApp usage, and responsiveness parameters are also depicted in Table 18. The results are as following:

- *Total Application Usage*: is positively related to median of challenge levels ($\tau = .32, p = .02, N = 31$). Total application usage is also in a positive relationship with the negative extreme response style of challenge ($\tau = .24, p = .02, N = 31$). It means that the participants who used mobile applications in a higher amount of time gave the response "not challenged at all" more than the ones who used mobile applications in a lower amount of time.

- *Total Social Category Usage*: is positively related to the negative extreme response style of engagement ($\tau = .27, p = .06, N = 24$) and challenge ($\tau = .27, p = .07, N = 24$). Those results can be inferred as the participants with higher amount of application usage in social category tend to select the item "not challenged/engaged at all" more than the participants with a lower amount of application usage in the social category.

- *Total Communication Category Usage*: is positively associated with the disacquiescence of challenge ($\tau = .26, p = .07, N = 24$) and negative extreme response style of challenge ($\tau = .34, p = .02, N = 24$). It can be stated that the participants who used communication applications more, recorded lower challenge scores than the ones who used communication applications less.

Table 18: Kendall's Tau correlation coefficients for the responsiveness variables, personality traits, and application usage parameters

		EXT	CONS	AGR	NEU	OPN	NV	TAU	TCOM	TSOC	TFB	TWA
Response	τ	-.18	-.02	.02	.07	-.38	.21	.19	-.19	.12	.17	.22
Rate	p	.29	.92	.92	.67	.03	.23	.13	.20	.40	.24	.14
Median of	τ	-.17	-.23	-.33	-.19	-.18	.22	.13	.07	.27	.15	.33
Engagement	p	.37	.22	.08	.31	.35	.24	.33	.66	.09	.34	.09
Interpolated Median	τ	-.08	.47	.12	.05	.15	.17	-.10	.09	.04	-.16	-.15
of Engagement	p	.62	.007	.48	.78	.40	.34	.43	.52	.80	.29	.32
Entropy of	τ	-.21	-.06	.01	.10	-.33	.33	.18	-.11	.03	.17	.08
Engagement	p	.22	.73	.94	.55	.06	.06	.16	.44	.82	.25	.57
Polarization of	τ	-.17	.30	.02	-.01	-.13	.31	.10	-.07	.20	10	.13
Engagement	p	.31	.08	.89	.97	.44	.07	.41	.66	.16	.50	.37
Acquiescence of	τ	-.27	.38	.12	-.01	-.02	.16	-.13	.19	.11	-.14	.01
Engagement	p	.11	.03	.50	.97	.89	.36	.31	.20	.46	.36	.96
Disacquiescence	τ	.10	-.29	-.09	-.10	-.12	-.17	.08	.03	.10	.26	.16
of Engagement	p	.55	.09	.62	.55	.48	.34	.55	.82	.50	.07	.26
Acquiescence Balance	τ	-.18	.40	.15	.06	.06	.16	-.13	.10	.02	-.18	-.08
of Engagement	p	.29	.02	.39	.73	.72	.36	.32	.49	.88	.22	.59
Middle RS	τ	.18	-.29	-.11	.13	.05	.04	.01	-.09	-.21	-.13	-.14
of Engagement	p	.31	.09	.52	.44	.78	.83	.95	.57	.15	.37	.33
Positive Extreme RS	τ	-.27	.40	.05	.09	-.04	.18	-.02	.11	.28	.02	.03
of Engagement	p	.13	.02	.77	.59	.80	.32	.86	.45	.06	.90	.84
Negative Extreme RS	τ	-.04	-.44	-.35	-.10	-.25	.09	.20	.06	.27	.37	.42
of Engagement	p	.81	.010	.04	.57	.15	.62	.11	.67	.06	.01	.004
Median of	τ	-.15	-.13	-.42	-.05	-.13	.49	.32	.01	.08	.07	.32
Challenge	p	.42	.46	.03	.80	.49	.009	.02	.94	.62	.68	.04
Interpolated Median	τ	.18	.35	.11	-.18	.06	.21	-.07	-.24	-.04	-.28	-.14
of Challenge	p	.29	.04	.55	.29	.75	.23	.57	.11	.78	.06	.36
Entropy of	τ	-.31	.14	.00	-.01	-.49	.34	.06	-.24	-.23	-.11	.02
Challenge	p	.07	.40	1.00	.97	.005	.05	.62	.10	.12	.44	.88
Polarization of	τ	-.15	.30	.04	-.04	-.19	.28	.07	-.08	.09	-.01	.04
Challenge	p	.38	.08	.83	.81	.26	.11	.57	.59	.52	.94	.80
Acquiescence of	τ	-.18	.33	.11	-.03	-.13	.21	-.11	.01	.12	-.10	.01
Challenge	p	.31	.06	.54	.86	.46	.23	.39	.96	.40	.50	.96
Disacquiescence	τ	-.08	-.35	-.08	.15	-.09	-.21	.05	.26	.06	.28	.19
of Challenge	p	.65	.04	.64	.38	.62	.24	.72	.07	.69	.06	.20
Acquiescence Balance	τ	-.04	.35	.07	-.06	-.03	.30	-.07	-.14	-.02	-.22	-.14
of Challenge	p	.83	.04	.67	.72	.86	.09	.57	.33	.90	.14	.33
Middle RS	τ	.03	.30	.20	-.10	.11	.07	-.08	-.32	-.10	-.12	-.21
of Challenge	p	.86	.08	.25	.55	.52	.70	.53	.03	.52	.41	.16
Positive Extreme RS	τ	-.31	.34	-.05	.23	-.30	.51	.09	.00	.00	-.17	.02
of Challenge	p	.09	.07	.79	.21	.10	.006	.51	1.00	1.00	.27	.92
Negative Extreme RS	τ	-.07	-.24	-.20	.19	.02	-.07	.24	.34	.27	.36	.31
of Challenge	p	.70	.16	.24	.27	.89	.70	.06	.02	.07	.02	.04

EXT: Extroversion, CONS: Conscientiousness, AGR: Agreeableness, NEU: Neuroticism, OPN: Openness, NV: Negative Valence, TAU: Total application usage, TCOM: Total communication category usage, TSOC: Total social category usage, TFB: Total Facebook usage, TWA: Total WhatsApp usage, RS: Response style.

Table 19: Kendall's Tau correlation coefficients for the personality traits and the application usage parameters

		Total App Use	Communication	Social	Facebook	WhatsApp
Extroversion	τ	-.02	-.08	.06	-.21	.04
	p	.92	.69	.77	.27	.84
Conscientiousness	τ	-.16	.02	-.14	-.27	-.12
	p	.36	.92	.49	.16	.55
Agreeableness	τ	.04	.06	-.16	-.16	-.38
	p	.83	.76	.42	.42	.06
Neuroticism	τ	.15	.18	-.04	-.08	.06
	p	.38	.37	.84	.69	.76
Openness	τ	.06	.04	.12	.08	-.24
	p	.72	.84	.55	.69	.23
Negative Valence	τ	.28	.05	-.13	-.17	.51
	p	.11	.80	.52	.39	.01

N=19 for the Total App Use parameter, N=15 for the others.

- *Total Facebook Usage*: is in a negative relation with the interpolated median of challenge ($\tau = -.28, p = .06, N = 24$) which means that the participants who used Facebook more tend to respond challenge questions with lower scores than the participants who used Facebook less. Similarly, total Facebook usage is in a positive relation with disacquiescence of engagement ($\tau = .26, p = .07, N = 24$), with disacquiescence of challenge ($\tau = .28, p = .06, N = 24$), with the negative extreme response style of engagement ($\tau = .37, p = .01, N = 24$), and with the negative extreme response style of challenge ($\tau = .36, p = .02, N = 24$). The results infer that the participants who used Facebook a higher amount of time recorded lower engagement/challenge scores than the participants who used Facebook lower amount of time.
- *Total WhatsApp Usage*: is in positive relation with the medians of engagement ($\tau = .33, p = .03, N = 24$) and challenge ($\tau = .32, p = .04, N = 24$). Similar to Facebook usage, WhatsApp usage is also in a positive relation with the negative response styles of engagement ($\tau = .42, p = .004, N = 24$) and challenge ($\tau = .31, p = .04, N = 24$). The results infer that the participants who used WhatsApp higher amount of time during work hours responded with "not challenged/engaged at all" item more than the ones who used WhatsApp lower amount of time.

Although it was not hypothesized, the correlation coefficients between BPTI scores and mobile application usage constructs were calculated. The results are presented in Table 19. WhatsApp usage during work hours is observed to be positively related with Negative Valence ($\tau = .51, p = .01, N = 15$), and negatively related with Agreeableness ($\tau = -.38, p = .06, N = 15$). It can be inferred that the participants who evaluate their personality more negatively, used WhatsApp during work hours less than the ones who have lower negative valence scores. On the other hand, the more agreeable participants used WhatsApp a lower amount of time than the ones who are less agreeable.

The research framework given in Figure 17 has been revised based on the relations found in the analyses as in Figure 19. The dashed lines show the hypotheses partially supported, the black lines show the hypotheses supported, and the red lines show the relations not hypothesized at first, but found significant.

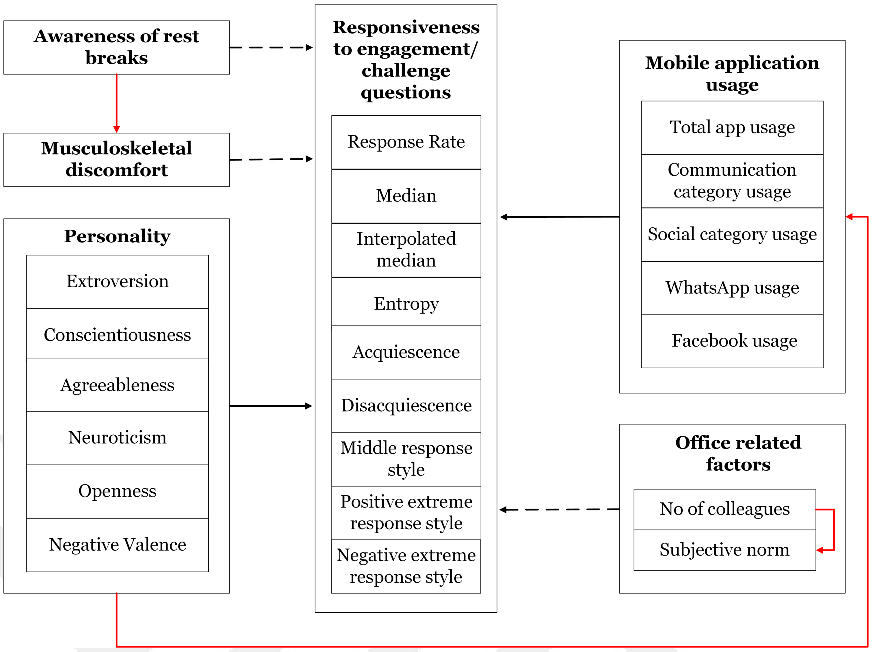


Figure 19: Revised research framework based on the analysis results

5.5 Discussion

The results of the study show three groups of variables (personality, mobile application usage, and office-related factors) had significant relations with the responsiveness of knowledge workers regarding their engagement/challenge levels. The remaining two groups of variables (awareness and musculoskeletal discomfort) had relations at a lower level. In the following, the relations of those groups are discussed in detail.

5.5.1 Musculoskeletal Discomfort and Awareness

Musculoskeletal discomfort was found positively related to the entropy of engagement responses. The engagement responses of the participants who felt musculoskeletal discomfort in higher levels were more heterogeneous than the ones who felt musculoskeletal discomfort in lower levels. Surely more evidence is needed, but a high degree of musculoskeletal discomfort could be the reason for not focusing on work, hence the responses were varied more.

The participants who are more aware of giving rest breaks responded with the item “I am not engaged at all” higher number of times than the ones who are less aware. The result may be due to the fact that since those participants are more aware of the importance of giving rest breaks, they may actually give rest breaks more frequently than the ones who are not aware that much. For this reason, the ESM messages may have arrived at the breaks, so that they may have responded as they were not engaged with their work at that moment. A significant relationship with other responsiveness variables were not found. As van Kenhove et al. [175] stated, response behaviour can be related to topic involvement, and involvement of individuals must exceed a critical level for the decision of participation. Similarly, in this study, the participants' awareness levels or musculoskeletal discomfort levels may not have exceeded a critical level for themselves, so their responsiveness metrics may not have been related to their awareness or musculoskeletal discomfort levels.

Although it was not hypothesized the relation between musculoskeletal discomfort and awareness, the results revealed that there is a negative relation between the two. The participants who take rest breaks more regularly stated that they do not feel musculoskeletal discomfort while working as much as the ones who do not take regular rest breaks. On the other hand, the participants who do not take regular rest breaks were the ones who feel musculoskeletal discomfort more. The negative relation between the degree of feeling musculoskeletal discomfort during work and the degree of taking rest breaks clearly showed that participants who take regular rest breaks from their work reported suffering less from musculoskeletal discomfort. This result showed the importance of the rest breaks as in previous studies [12, 13, 17].

5.5.2 Office-related Factors

The number of colleagues and the subjective norm scores correspond to the office related factors in the study. Based on the results, the number of colleagues was found in a weak positive association with the entropy of challenge responses. Previous studies mostly focused on office types and their effects on distractions [32, 123]. As those studies stated, employees in shared offices stated that their attention distracts more easily than those in private offices. The results contribute to those results by presenting the number of colleagues in the same office. The number of colleagues that is included in this study might be a strong indicator of office type. Although office type was not measured directly, it might be inferred based on the number of colleagues. For example; the maximum number of colleagues stated in the questionnaire was 50, which is surely an indicator for a shared office. So, it might said that employees in the shared offices responded with a variety of challenge items. It may be due to the distraction they perceive in their office environments. As suggested by the previous studies, distraction in shared offices is higher, that may lead to different challenge levels of employees in such offices.

Again, even though it was not hypothesized, the positive relationship between the number of colleagues and the subjective norm scores is found statistically significant. The participants who shared their offices with a higher number of colleagues stated that they affect more from their colleagues for taking rest breaks compared to the ones who share a lower number of colleagues. The result is in line with the previous studies, still, to the best of our knowledge, the relation found in the study has not been stated in any study before. Male employees in a university work environment stated that participating in a social group is a motivating factor [31]. It is contributed to this result by showing a significant direct relation between the size of that social group and the degree of the social norm. The bigger the social group is, the higher the people are affected by that social group.

5.5.3 Personality

The results of this study revealed significant relationships among personality traits, responsiveness, and engagement/challenge levels.

- *Extroversion*: Based on the results of this study, extrovert participants have more homogeneous challenge responses than introverts. Besides, introverts tend to respond with "I am very challenged" item more likely than extroverts. When two results are combined, it can be said that extroverts tend to respond more homogeneously, but mostly they responded with lower levels of challenge. On the other hand, introverts responded with higher number of categories; still, their likelihood of responding with positive extreme category (which is "very challenged") is higher than the extroverts. Previous work found positive relations with Extroversion and extreme response styles [80–83]. However, in this study, the positive extreme response style is negatively related to the Extroversion scores. It may be due to the differences in the application domain where the responses are measured. Extroversion is related to being confident, risk-taking, and proactive; so extrovert participants may not have felt challenged by their work most of the time. This may have caused a decrease in positive extreme responses for challenge questions. In this point of view, the results contribute to the literature.

- *Conscientiousness*: is the most related trait to work engagement [111–114]. Similarly, in this study, it is found significantly related to both engagement and challenge levels, and responsiveness metrics. First of all, conscientious individuals are responsible, well-organized, hard-working [132, 133]. Hence, it is expected them to be more engaged with their work during work hours. As expected, the results showed that conscientious participants responded with the higher levels of both engagement and challenge. They tend to give positive responses a higher number of times (i.e. they have more positive extreme response style regarding engagement and challenge levels) than the ones with low conscientiousness scores. Similarly, they have less negative extreme response style regarding engagement and challenge levels. Previous studies showed positive relations with Conscientiousness and extreme response style [81,83]. Hibbing et al. [80] also showed the positive relation between conscientiousness and acquiescence response style. In this study, a positive relation between Conscientiousness and acquiescence response style was found, and this result was contributed by adding a negative relation with Conscientiousness and disacquiescence response style.

- *Agreeableness*: is negatively related to the median of engagement and challenge levels. Agreeable individuals are known as good-natured, cooperative, and trustful [132] whereas the opposite is known as uncooperative, stubborn, and sceptical [133]. The reason for the agreeable participants faced lower challenge levels might be due to their cooperative manners so that they did not get challenged with their work as much as the degree of less agreeable participants. The results also indicate that there is a negative relation with Agreeableness and negative extreme response style for engagement. None of the previous studies but except the one of He and van de Vijver [81] found significant relations between Agreeableness and extreme response style. Hibbing et al. [80] criticised that result since the nature of Agreeableness avoids from "extreme" situations. This study contributes to that result. It was showed that agreeable participants avoided from giving negative extreme responses to the engagement questions, and this is in line with their personalities.

- *Openness*: is the only trait that is directly related to the response rates of the participants. The participants who are high on Openness gave a lower number of responses to the overall ESM questions. Previous studies show similar results: high open individuals were more likely to refuse to participate in surveys [176], or they were the late-semester participants [177, 178]. Similarly, the low response rate may have been due to the nature of individuals who are high on Openness: they are receptive to new ideas and experiences. The ESM questions were sent six times a day in this study so they may have got bored of answering the same questions, and this may have resulted in low response rates of such individuals. On the other hand, when participants with high Openness scores respond, they tend to give more homogeneous responses.

- *Negative Valence*: The participants who perceive themselves more negatively recorded higher levels of engagement and challenge levels than the ones who perceive themselves more negatively. On the other hand, the participants with high scores on Negative Valence had a higher number of positive extreme response of challenge. Negative manner (i.e. high negative valence) might affect the perception of work challenge in a negative way, i.e., such personalities may have felt more overwhelmed with work than individuals who were more easy-going and in positive perspective to themselves.

Previous studies mostly focused on the relationship between work engagement and personality. Specifically, Extroversion, Neuroticism and Conscientiousness were found related to work engagement [113, 114]. This study contributes to those studies with the investigation of personality and work challenge levels. Several metrics of responsiveness were also included in this study and they were also found related to the personality.

In addition to those relations, the results showed a significant negative relation between Negative Valence and WhatsApp usage. Participants who evaluate his/her personality more negatively used a lower amount of WhatsApp during work hours than the ones who have lower negative valence scores. Previous studies about personality and WhatsApp usage found significant relations between the Big Five personality traits and WhatsApp usage [179]. To the best of our knowledge, no study has worked with BPTI and the sixth trait named Negative Valence. This study contributes to the literature by showing the relatedness of Negative Valence and WhatsApp usage by focusing the WhatsApp usage only in work hours.

5.5.4 Application Usage

The results show that there are significant relations between the responsiveness of participants and application usage parameters. The participants who used mobile applications a higher amount of time during work hours tend to respond to the engagement/challenge questions with lower levels than the ones who used mobile applications lower amount of time. When the results are investigated in metric-level, it can be seen that disacquiescence or negative extreme response styles of engagement/challenge are in positive relationships with almost all application usage parameters. So, an increase in application usage during work hours may be a signal for a low level of engagement/challenge with work. The studies [2, 111] showed that Facebook usage (on web browsers) is significantly effective on engagement/challenge levels. It is contributed to this result by considering Facebook usage as mobile application usage.

The results show that there is a significant positive relation between total application usage and the median of challenge levels. However, no significant relation was found between interpolated median of engagement/challenge levels and total application usage. Although the relationships are not significant, the direction of the relations was negative. Hence, both results are not consistent. Recall that using median especially for Likert scales has a limitation, therefore interpolated median was used in the study. Besides, those aggregated measures are limited on analyzing the relations at individual level. Every user has a different level of application usage, therefore it is more meaningful to analyze those relationships at individual level. An increase or a decrease at individual level could be seen directly with such analyses. Hence, repeated-measures results could be more reliable in this circumstance. Thus, the relations are investigated with repeated-measures analyses in the next chapter.

CHAPTER 6

MODELLING IN-SITU ATTENTIONAL STATES, ENGAGEMENT AND CHALLENGE LEVELS OF OFFICE WORKERS

In this chapter, the effects of mobile phone usage on the inference of *in-situ* work engagement/challenge levels and attentional states of knowledge workers are investigated. A brief introduction is presented in the following section. Then, the method used in the study is explained. The results are presented and discussed in the last sections.

6.1 Introduction

Attentional states, as well as engagement and challenge levels of office workers, can indirectly be inferred via mobile phones. For example, an increase in smartphone usage may be a sign of boredom since most users prefer to use mobile phones when they feel bored [20–22]. Similarly, their interaction with computers (which applications they use and how long) might also reveal boredom [2, 23]. Besides, there is a need for personalization for making inferences about attentional states, and engagement/challenge levels. The relative differences among office workers need to be considered in the models developed. However, at the early stages of the design of personalized models, the cold-start problem may occur since the number of data points is relatively small as stated before.

In the study presented in this chapter, it is investigated which application usage metrics are effective on inferring *in-situ* work engagement/challenge levels and attentional states. The model is also offered as a solution to the cold-start problem. Basically, the study is conducted to investigate the research questions:

- *Which application usage metrics are related to in-situ engagement/challenge levels of office workers?*

- *How can a model be built for inferring attentional states and engagement/challenge levels of office workers using application usage metrics by considering cold start problem, the variety in the number and characteristics of the responses, and repeated-measures design of the data? How is this model comparable to individual and general models?*

6.2 Method

In this study, the method explained in the previous section was employed (see Figure 18). The ESM responses from 14 participants from whom the application package names were obtained, were used. The details of the data set is given in next section. Repeated-measures correlation was used for investigating the relation between *in-situ* engagement/challenge levels and application usage parameters. Then, Generalized Linear Mixed Model was employed for inferring the attentional states and engagement/challenge levels, and the model performance was compared with general and individual random forest models, which are commonly used in the literature.

6.3 Feature Set

As remembered, the number of sessions recorded from 14 participants (those whose application package names could be obtained and had a 25% or higher response rate in ESM questions) is 418 in total. The features regarding the application usage are given below. Each feature is calculated for each participant in the last 5, 10, 15, 30, 45 or 60 minutes separately before each response to ESM delivery:

- *Application usage (AU)*: refers to the total duration of application usage (in minutes). For example, AU_{10} refers to the last 10-minute usage.
- *Number of applications (NOA)*: indicates the number of unique applications used.
- *Number of switches (NOS)*: is the total number of transitions between mobile applications.
- *Mean application usage (MAU)*: corresponds to the average duration of application usage (in minutes).
- *Communication category usage (COM)*: denotes the duration of application usage in the communication category.
- *Social category usage (SOC)*: denotes the duration of application usage in the social category.
- *Facebook usage (FB)*: is the duration of usage in Facebook application. Facebook was selected from the social category since Facebook usage was higher than half of the social category usage (57.40%).
- *WhatsApp usage (WA)*: shows the duration of usage in WhatsApp application. WhatsApp was selected from the communication category since WhatsApp usage was nearly equal to half of the social category usage (47.79%).
- *Messaging applications usage (MES)*: indicates the duration of usage in messaging applications (namely WhatsApp, Facebook messenger and SMS usages) belong to communication category. The usage of messaging applications was equal to 58.54% of the communication category usage.

ESM responses consisted of engagement and challenge levels recorded with each ESM questionnaire, and attentional state derived from engagement and challenge scores for the same ESM questionnaire. The attentional states of the participants were classified using challenge and engagement levels of

Table 20: The descriptive statistics of the engagement/challenge responses and attentional states obtained from ESM questionnaires.

Parameter Name	Values	Frequency	Percentage
Engagement	1	132	31.58%
	2	73	17.46%
	3	52	12.44%
	4	66	15.79%
	5	95	22.78%
Challenge	1	146	34.93%
	2	89	21.29%
	3	81	19.38%
	4	60	14.35%
	5	42	10.05%
Attentional states	Bored	197	47.13%
	Focused	175	41.87%
	Rote	38	9.09%
	Frustrated	8	1.91%

participants as in [2]. For example; if a participant recorded 1 as engagement response and 1 as challenge response, then attentional state of the participant at that moment was labelled as "bored". The descriptive statistics of engagement and challenge responses obtained from ESM questions are given in Table 20. Only "focused" and "bored" states were included in the analyses since the number of data points in "rote" and "frustrated" states had a fewer number of data points.

6.4 Repeated-Measures Correlation Results

In this section, it is investigated *which application usage metrics are related to in-situ engagement/challenge levels of office workers*. The repeated-measures correlation (rmcorr) coefficients were calculated on the ESM answer data set between in-situ engagement/challenge levels and the application usage constructs: *application usage, number of applications, number of switches, mean application usage, social applications, communication applications, messaging applications, WhatsApp* and *Facebook usage* in order to measure the relation between in-situ engagement/challenge levels and mobile application usage. Rmcorr results between application usage constructs and in-situ engagement levels are given in Table 21. Similarly, rmcorr results between application usage constructs and in-situ challenge levels are given in Table 22. Note that as the attentional state is a binary variable, rmcorr could not be applied.

The results show that the number of switches and the number of applications are the most related features to challenge levels in decreasing order. In particular, the window size between 30 and 60 minutes appears to be the most determinant one in common. Although the window size between 10 to 30 appears to be statistically significant, the magnitude of the relations appears to be lower than the ones having a longer time window. All the relations are in a negative direction. For example, in 60 minute window size, as the number of switches increases, engagement ($r_{rm} = -.11, p = .02$)

Table 21: Repeated measures correlation results between application usage variables and engagement

		5-min	10-min	15-min	30-min	45-min	60-min
App Use	r_{rm}	-.14	-.12	-.08	-.05	-.04	-.04
	p	.005	.02	.12	.28	.37	.45
No of Apps	r_{rm}	-.11	-.13	-.08	-.11	-.08	-.09
	p	.02	.009	.11	.03	.11	.08
No of Switch	r_{rm}	-.11	-.11	-.07	-.09	-.12	-.11
	p	.03	.02	.17	.08	.02	.02
Mean App Use	r_{rm}	-.06	-.07	-.05	-.01	.01	.04
	p	.24	.18	.28	.80	.91	.37
Social	r_{rm}	-.04	-.05	-.01	.00	.01	.03
	p	.46	.30	.81	.97	.79	.58
Communication	r_{rm}	-.10	-.11	-.05	-.05	-.07	-.08
	p	.05	.03	.28	.35	.13	.10
Messaging	r_{rm}	-.07	-.09	-.06	-.07	-.07	-.06
	p	.14	.08	.25	.15	.14	.26
Facebook	r_{rm}	-.08	-.05	-.02	.06	.07	.04
	p	.11	.46	.75	.21	.14	.44
WhatsApp	r_{rm}	-.03	-.03	-.01	-.04	-.05	-.04
	p	.54	.53	.85	.38	.35	.48

Rows show application variables, columns show the time window of the variables. ($N = 418$)

and challenge levels ($r_{rm}, p < .001$) decrease, or vice versa. Similarly, the number of applications in window size 60 minutes has a negative correlation with engagement ($r_{rm} = -.09, p = .08$) and challenge levels ($r_{rm} = -.14, p = .006$). In other words, it means that application usage in a higher amount of time refers to a lower level of engagement/challenge levels. In terms of the application category types, the use of communication applications is significantly related to the challenge levels (for 60-minute window size ($r_{rm} = -.10, p = .04$)) but not social category type applications.

The total application usage, number of applications, number of switches and communication type applications are the most related features to engagement levels in decreasing order. Similarly, all the relations are also in a negative direction. However, this time, the window size between 5 and 10 appear to be the most determinant one in common apart from the fact that the number of switches appears to be also significant in longer time windows. The use of social, Facebook, WhatsApp and messaging applications were not found as significantly related to engagement and challenge levels.

6.5 GLMM for In-Situ Attentional States, Engagement and Challenge Levels

In this section, *a model for inferring attentional states, engagement/challenge levels of office workers using application usage metrics by considering cold start problem, the variety in the number and characteristics of the responses and repeated-measurement nature of the data* is developed. It is also

Table 22: Repeated measures correlation results between the application usage variables and challenge

		5-min	10-min	15-min	30-min	45-min	60-min
App Use	r_{rm}	-.11	-.09	-.10	-.09	-.10	-.09
	p	.03	.08	.06	.06	.04	.08
No of Apps	r_{rm}	-.09	-.12	-.11	-.17	-.13	-.14
	p	.07	.02	.03	<.001	.008	.006
No of Switch	r_{rm}	-.10	-.11	-.10	-.17	-.18	-.17
	p	.05	.03	.04	<.001	<.001	<.001
Mean App Use	r_{rm}	-.02	-.04	-.05	-.04	-.03	.00
	p	.63	.38	.28	.43	.55	.92
Social	r_{rm}	-.01	-.04	-.04	-.05	-.04	-.03
	p	.87	.45	.38	.29	.37	.51
Communication	r_{rm}	-.09	-.10	-.08	-.07	-.11	-.10
	p	.06	.04	.13	.18	.03	.04
Messaging	r_{rm}	-.05	-.07	-.05	-.07	-.07	-.05
	p	.28	.17	.27	.18	.17	.29
Facebook	r_{rm}	-.05	-.03	-.04	-.02	-.02	-.06
	p	.28	.48	.45	.66	.70	.25
WhatsApp	r_{rm}	-.04	-.04	-.03	-.05	-.05	-.04
	p	.44	.37	.48	.28	.28	.44

Rows show application variables, columns show the time window of the variables. ($N = 418$)

investigated how the model is comparable to individual and general models, which use the random forest method.

6.5.1 Model and Feature Set Selection using GLMM

The repeated-measures correlation results gave an idea about which variables of application usage are related to engagement/challenge levels. However, it is a need to investigate which of these features are indeed effective for modelling. As rmcorr shows the linear association between the variables and GLMM is inherently a linear model, the correlation results were taken as a basis in feature set selection and model building instead of variable selection with Gini or other metrics.

As can be seen from the Tables 21 and 22; AU_5, NOA_5, NOA_{10}, NOA_{30}, NOA_{45}, NOA_{60}, NOS_5, NOS_{10}, NOS_{30}, NOS_{45}, NOS_{60}, COM_5, COM_{10}, COM_{45} and COM_{60} are the most related variables to the engagement and challenge levels. Hence, it was planned to include a combination of those variables in the GLMM analysis. However, before fitting a GLMM, the correlations between the predictor variables were calculated in order to detect possible multicollinearity issues. Strong correlations were observed between some of the variables such as NOS_{45} and NOA_{45} ($r_{rm} = .91, p < .001, N = 418$) signalling a multicollinearity problem. Later, the pairs which are NOS_{60}, NOA_{10}, AU_5 and NOS_{45} having a correlation less than .70 were taken into account. The combination of those variables were given to the GLMM trials.

Then, several GLMMs were built for modelling attentional states, engagement levels and challenge levels of users. During this step, a different approach from the studies in the literature was taken. The data set was sub-sampled five times rather than solely using the original one. This is due to the fact that the model and feature selection should not be affected by the number of responses or the highest/lowest amount of application usage. If the majority of data comes from a few users in the data set, selected features and model might be not representative for all users. In the sub-sampling approach, in each sub-sample, the data points of the participants with the highest and lowest response rates were eliminated incrementally. To be more specific, the first sample is the full original data set consisting of all the responses of users. In the second data set, two users were removed from the first data set. These users were the ones with the highest and lowest number of ESM responses. In the third data set, two more users with the highest and lowest number of ESM responses were excluded from the second data set. In each iteration, a more uniform data set in terms of participants' responses than the data set from the previous step, was obtained. At the end of this process, the model, which gave consistently the lowest DIC on all the data sets, was selected. The lowest DIC is the preferred metric for Bayesian model selection [150].

The response variables are binary (attentional states) and ordinal (engagement and challenge levels). The model fitting process was carried out incrementally by adding constructs one at a time according to their geometric mean performance across the data sets. Each model was run five times. In the end, the mean and standard deviation of the overall performance of each model was reported.

For modelling attentional states, first, NOS_{60}, NOA_{10}, AU_5 and NOS_{45} variables were given to GLMMs separately. Then, those variables were given to the model by pairs. In total, 11 combinations of those variables were used for predicting attentional states. Appendix D.1 presents the model runs for predicting attentional states on five different data sets with the mean and standard deviation of the DIC values for each run. Similarly, engagement and challenge levels of users were modelled with GLMM. Same 11 models for each were built. Appendix D.2 and Appendix D.3 summarize the DIC values obtained for the models of engagement and challenge models, respectively. For all target variables (attentional states, engagement and challenge levels), the models with NOS_{45} and AU_5 predictors gave the lowest DIC; hence, those models were selected.

6.5.2 GLMM Results

Table 23 (top) shows the posterior distributions of each parameter with posterior means and 95% credible intervals (2.5 and 97.5 percentiles of the posterior distribution) of Model 3 for attentional states. The number of switches in the last 45 minutes (NOS_{45}) has been found statistically significant $p = .02$) on predicting attentional states. The negative relation between NOS_{45} and attentional states means that as the number of switches in the last 45 minutes before ESM messages increased, users were more likely to be "bored"; or as *NOS* decreased, users were more likely to be "focused". Similarly, application usage in the last 5 minutes before ESM messages has been found negatively related to the attentional states. As the duration of application usage in the last 5 minutes increased, users were more likely to be in the "bored" state, or vice versa. Based on the magnitudes, it can be said that the effect of application usage in the last 5 minutes is higher than the number of switches in the last 45 minutes.

Table 23 (middle) shows the posterior distributions of each parameter with posterior means and 95% credible intervals (2.5 and 97.5 percentiles of the posterior distribution) of Model 1 for engagement levels. Similar to the attentional states model, the number of switches in the last 45 minutes and the duration of application usage in the last 5 minutes are negatively related to the engagement levels. As the number of switches increased, participants tended to be lowly engaged with their work or vice versa. Similarly, as participants used a higher amount of mobile applications in the last 5 minutes, they were more likely to be less engaged with their work or vice versa. As in the attentional states model, the magnitude of application usage is higher than the number of switches on the engagement levels.

Finally, Table 23 (bottom) shows the posterior distributions of each parameter with posterior means and 95% credible intervals (2.5 and 97.5 percentiles of the posterior distribution) of Model 3 for challenge levels of users. Again, the effect of NOS_{45} parameter was significantly negative on challenge levels ($p = .008$). As *NOS* in the last 45 minutes increased, challenge levels of users decreased, or vice versa. Similarly, application usage in the last 5 minutes is also in a significant negative relation with the challenge levels ($p = .01$). The participants who used mobile applications a higher amount of time in the last 5 minutes tended to be low challenged with their work or vice versa.

Table 23: Posterior means, 95% credible intervals and p values of parameters for Model 3 for attentional states (top), Model 1 for engagement levels (middle) and Model 3 for challenge levels (bottom)

Model 3 for Attentional States			
Parameters	Posterior mean	95% CI	p
(Intercept)	.23	(-.17,.60)	.21
NOS_{45}	-.07	(-.12,-.01)	.02
AU_5	-.38	(-.73,-.10)	.01
Model 1 for Engagement Levels			
Parameters	Posterior mean	95% CI	p
(Intercept)	.92	(.68,1.11)	<.001
NOS_{45}	-.04	(-.07,-.01)	.008
AU_5	-.21	(-.39,-.04)	.02
Model 3 for Challenge Levels			
Parameters	Posterior mean	95% CI	p
(Intercept)	.84	(.62,1.09)	<.001
NOS_{45}	-.06	(-.10,-.02)	.002
AU_5	-.15	(-.31,.03)	.01

6.5.3 Comparison Results

As stated before, for the comparison of GLMM, random forest models were fit both in general (population)-level and individual-level. The variables used in the models were the same with the ones reported in GLMM results: number of switches in the last 45 minutes (NOS_{45}), and application usage in the last 5 minutes (AU_5) before ESM messages. The random forest models were built for each participant (in total 14 models), then the accuracy values were reported by averaging them.

The repeated random sub-sampling validation (i.e. repeated hold-out) [162] was used for building models for all the classifiers. The data set was divided into training and test sets with the proportions of 30-70, 40-60, 50-50, 60-40, and 70-30, and repeated each 20 times. When dividing data into training and test sets, stratified sampling, which enables to balance class proportions in each set, was used. The accuracy of the models is reported as the performance metric. The estimated accuracy is obtained by averaging 20 runs. The only difference for the individual random forest classifier is that the data set consisted of one user's responses.

For all random forest classifiers (general and individual); as the parameter of the number of trees, several values (50, 100, 150, 250, 500, 750, 1000) were attempted in the models with cross-validation on the data set, which was reserved as training data set in the experiments. The optimized parameters were found as 750 for the general attentional states model, 50 for the individual attentional states model, 100 for the general engagement levels model, 200 for the general challenge levels model, and 500 for the individual engagement and challenge levels model. The baseline performance was also calculated with the majority classifier that always predicts the class with the highest number of data points.

In Table 24, the average accuracy values obtained from four different classifiers for predicting attentional states, engagement levels, and challenge levels are reported with their standard deviations. In the second column, the percentage of the training set is stated. Remember that attentional states were modelled as a binary response (as "focused" and "bored"), engagement and challenge levels were modelled as an ordinal response (as 1-5).

As illustrated in the table, GLMM predicted engagement and challenge levels better than the general random forest model, individual random forest and baseline classifiers. Only for predicting attentional states, individual random forest model was slightly better than GLMM. In order to compare the accuracy values of four classifiers, statistical tests were conducted on the accuracy values obtained from all runs (5 different training percentages x 20 runs = 100 accuracy values for each classifier). The Shapiro-Wilk Test showed that the accuracy values of the models did not distribute normally ($p < .001$). Hence, the Friedman Test was performed [166]. The results of Friedman Test showed that the average accuracy values obtained from four classifiers are significantly different for the prediction of attentional states ($\chi^2(3) = 169.07, p < .001, N = 100$), engagement levels ($\chi^2(3) = 206.06, p < .001, N = 100$), and challenge levels ($\chi^2(3) = 208.50, p < .001, N = 100$). Then, Wilcoxon signed rank tests were conducted for binary comparisons of the models as post-hoc tests. For the prediction of attentional states, the accuracy obtained from GLMM is significantly higher than the general random forest's accuracy ($Z = -8.26, p < .001, N = 100$), individual random forest's accuracy ($Z = -5.99, p < .001, N = 100$), and the baseline accuracy ($Z = -4.66, p < .001, N = 100$). Similarly, GLMM significantly outperforms general random forest ($Z = -8.68, p < .001, N = 100$), and individual random forest ($Z = -8.59, p < .001, N = 100$) for predicting engagement levels. The difference between GLMM accuracy values and baseline accuracy values is significantly different at the α=.1-level ($Z = -1.87, p = .06, N = 100$). Finally, for the challenge levels models, GLMM gives significantly the most accurate results among general random forest ($Z = -8.68, p < .001, N = 100$), and individual random forest ($Z = -8.02, p < .001, N = 100$). However, the difference between GLMM and the baseline classifier is not found statistically significant ($Z = -.52, p = .60, N = 100$), which means that baseline classifier and GLMM give similar accurate results for predicting challenge levels.

Table 24: Comparison of model accuracy values for predicting attentional states, engagement and challenge levels

Model	Training Percentage	Attentional States $N = 372$	Engagement $N = 418$	Challenge $N = 418$
	30%	53.53%±3.28%	29.35%±3.29%	33.23%±2.82%
	40%	54.02%±2.09%	31.97%±1.92%	34.17%±3.01%
GLMM	50%	53.22%±3.45%	32.03%±2.28%	36.10%±2.05%
	60%	55.51%±2.96%	31.62%±2.94%	36.31%±3.09%
	70%	54.19%±3.17%	34.12%±3.73%	35.97%±2.91%
General RF	30%	48.34%±2.31%	21.51%±2.73%	26.56%±3.21%
(No of trees =	40%	48.26%±2.75%	22.61%±2.94%	28.89%±3.11%
750 for AS	50%	48.57%±2.91%	24.13%±2.40%	31.27%±2.48%
100 for Eng	60%	47.03%±2.95%	22.63%±1.40%	32.68%±2.69%
200 for Chal)	70%	49.10%±2.89%	23.08%±3.91%	31.98%±3.60%
Individual RF	30%	50.45%±12.95%	23.96%±11.87%	27.07%±13.87%
(No of trees =	40%	51.52%±15.32%	24.51%±12.86%	28.81%±14.38%
50 for AS	50%	49.73%±18.65%	24.17%±13.11%	30.57%±14.28%
500 for Eng	60%	52.18%±18.44%	24.54%±13.32%	29.43%±15.58%
500 for Chal)	70%	49.91%±21.98%	26.40%±16.94%	29.88%±17.71%
	30%	50.41%±18.26%	32.49%±15.22%	32.95%±14.54%
	40%	51.26%±18.28%	30.12%±13.92%	31.57%±14.47%
Baseline	50%	52.67%±19.89%	31.14%±15.71%	30.63%±14.71%
	60%	50.34%±19.67%	30.87%±18.40%	30.92%±16.61%
	70%	51.59%±22.74%	30.34%±21.60%	34.44%±18.02%

AS: Attentional States, Eng.: Engagement, Chal.: Challenge

6.5.4 Individual Models

As presented in Table 24, GLMM predicted attentional states, engagement and challenge levels better than individual random forest classifier. For a closer inspection of the individual models, each user's model accuracy values with their averages are reported in Table 25 for the individual random forest, GLMM, and baseline classifier. The results obtained with 20 runs on the 70% training data. The values illustrated in bold show the highest accuracy for each model (AS, engagement and challenge). The results are presented based on the participants' number of responses (N) in descending order. The participant's results whose data points are less than 20 are not reported since individual models with such a limited number of data points are not able to learn each response category characteristics, and also they are highly affected by the unbalanced categories.

It was also investigated whether individual model accuracy has a relation with the polarization or other responsiveness metrics mentioned in Chapter 5. In order to compare the individual models' accuracy values and the responsiveness indicators, a user table given in Appendix D.4 was prepared. The histograms of the engagement and challenge responses are also provided in the table. However, no relation was found between the metrics and the individual model accuracy values. The findings from the individual level models can be summarized as follows:

- Individual random forest classifier was not able to reach the accuracy values of GLMM and baseline most of the time.
- For attentional states models, the best accuracy was obtained with GLMM for nearly the half of the users, and with the baseline for the other half. The main reason for baseline classifier to be successful on the data set, its chance on being successful is 50% since attentional states are a binary variable. It can be seen from the table, as the number of data points decreases, the baseline classifier predicts the best among all. The performance of the baseline classifier increased up to 90% for U02. This is due to the low number of data points (N = 15), besides, the bias on the responses of U02. As can be seen from the table in Appendix D.4, the user responded engagement and challenge questions mostly with 1 or 2, which causes the attentional state for the user "bored" most of the time. Because of that, 90% of the time, the baseline classifier predicted the attentional state of U02 correctly, whereas other complex methods (GLMM and random forest) failed.
- On the other hand, GLMM mostly gave the highest accuracy values for engagement and challenge models. Since the range of engagement and challenge responses are wider (5 categories) than attentional states, baseline classifier became successful only when a user had a tendency to respond with the same engagement or challenge level item. For example; the challenge levels of U03 were predicted the most accurately with the baseline classifier. 44.4% of the challenge answers of that user consisted of item 2; hence, baseline classifier mostly predicted the true challenge level. GLMM or individual random forest classifier failed when users selected a specific item as their response most of the time. A similar situation also happened to the engagement levels of U08. That user mostly answered engagement questions with item 4, and that caused the baseline classifier predicted it correctly, but because of the low number of data points for the user (N = 13), other methods failed.
- Although it is not reported in the table, the standard deviations for baseline classifier are higher than GLMM or individual random forest classifier. It means that by chance, baseline classifier obtained higher accuracy on some folds, at the same time, it obtained very low accuracy on other folds. For example; the standard deviation of accuracy values for the baseline classifier on U03's challenge levels is 14.67% whereas GLMM has a standard deviation of 7.95% for the same user. It clearly shows that GLMM has more stable predictions than baseline classifier. Statistical tests also support that result.
- Although the result of Wilcoxon Test is not statistically significant for the comparison of GLMM and baseline classifier for challenge level models when the values are inspected at user-level, it can be seen that GLMM is a more appropriate method to predict challenge levels of the users.

6.6 Discussion

In this section, the results obtained from the repeated-measures correlation analysis and the generalized linear mixed model are discussed in detail.

Table 25: The average accuracy values for predicting attentional states, engagement and challenge levels with GLMM, individual random forest classifier, and baseline classifier

		Attentional States Models			Engagement Models			Challenge Models		
User No	N	GLMM	Ind. RF	Base	GLMM	Ind. RF	Base	GLMM	Ind. RF	Base
U13	50	**54.17%**	43.61%	43.61%	**28.00%**	20.67%	20.25%	**38.33%**	31.75%	37.00%
U14	49	**63.68%**	58.95%	43.68%	**40.00%**	24.00%	38.00%	**41.00%**	25.75%	40.00%
U03	45	**62.50%**	48.21%	51.43%	**33.21%**	29.44%	21.94%	35.83%	36.07%	**45.28%**
U06	37	**64.00%**	53.00%	43.67%	**56.82%**	35.00%	20.67%	**47.73%**	23.18%	32.67%
U10	36	59.55%	77.27%	**87.73%**	**34.55%**	28.21%	27.86%	60.71%	52.00%	**61.07%**
U07	32	55.50%	**61.00%**	54.50%	**40.50%**	14.00%	21.54%	**40.00%**	17.50%	37.69%
U01	31	53.50%	58.50%	**74.00%**	**40.00%**	23.75%	33.75%	**45.56%**	30.80%	35.42%
U04	31	55.42%	46.67%	**61.25%**	**41.67%**	25.00%	15.00%	**33.89%**	23.89%	21.25%
U11	27	47.00%	46.50%	**56.00%**	**28.75%**	14.38%	22.27%	33.75%	**38.75%**	36.50%
U05	24	49.38%	47.50%	**50.63%**	**30.71%**	14.29%	25.50%	**35.00%**	19.29%	20.00%
U02	15	45.00%	78.33%	**90.00%**	33.75%	**38.33%**	37.50%	**45.00%**	41.25%	39.00%
U09	14	**41.25%**	33.75%	38.75%	**35.00%**	13.75%	30.83%	**46.25%**	15.00%	30.00%
U12	14	33.33%	34.17%	**37.50%**	**57.50%**	36.25%	52.00%	**58.75%**	32.00%	40.00%
U08	13	40.00%	43.00%	**70.00%**	10.00%	15.00%	**65.00%**	**32.00%**	18.75%	26.00%

6.6.1 The Relation between Application Usage and In-Situ Engagement and Challenge Levels

As discussed in the previous chapter, aggregated measures have given an idea about the relation between the application usage and engagement/challenge levels. The median and the interpolated median of engagement and challenge levels were used in the study of the previous chapter. In this chapter, it has been investigated whether there is a relation between *in-situ engagement/challenge levels and application usage* with the consideration of *repeated-measures design*.

Repeated-measures correlation results showed a significant negative relation between *in-situ* engagement/challenge levels and total application usage. As total usage increases, participants' engagement and challenge levels decrease or vice versa. Similar relation occurred with communication applications usage. As participants' usage of communication applications increase, their engagement levels decrease or vice versa. In addition, the number of switches had a negative relation with engagement levels. The more participants switched between applications, the less they were engaged with their work. Similarly, as the number of applications used in the last 5 or 10 minutes increased, the engagement and challenge levels decreased. The results of the study are in line with the previous studies that investigated the relation between application usage and boredom [21, 22, 35, 119].

In this study, the relation between work engagement/challenge levels and application usage has been shown both in a short period of time (e.g. 5, 10 and 15 minutes) and in a longer period of time (e.g. 30, 45, 60 minutes). In the previous studies, different periods of time have been discussed. For example; in [74], they set time window as 60 minutes, whereas in [73] it was set as 10 minutes, and finally in [22] it was 5 minutes. Also, note that those studies did not investigate the work engagement, instead, they mainly focused on inferring user engagement with mobile phones or detecting boredom, however, they are worth to mention since they used similar application usage variables. This study shows the effects of different time windows of application usage. More specifically, in the short period of time, only communication category, total application usage, number of applications, and number of switches between applications have been found related to engagement/challenge levels.

6.6.2 GLMM and Individual Models

Based on the results of GLMM that was fit for modelling attentional states, in-situ engagement/challenge levels, NOS_{45} and AU_5 are negatively related to the attentional states, engagement, and challenge levels. As NOS_{45} and AU_5 increased, the participants were most likely to be in "bored" state, i.e., work their engagement and challenge levels decreased. It is in line with the previous studies. Similarly, previous studies $[21, 22, 73]$ showed that an increase in application usage is a sign of boredom. As stated above, the time interval used in those studies differed from the settings of this study. When the time intervals are considered, it can be inferred that the number of switches in a longer time is effective, whereas duration of application usage is determinant in a shorter period of time. Since boredom is a state of mind in which one searches for a stimulus, in today's world, most of the mobile phone users engage with their devices when they are in such a mood. As expected, when users switch between mobile applications, it may be a sign of boredom, and that means the user is not engaged or challenged with his/her work. Similarly, when users seek a stimulus, they used their mobile phones for a longer duration. Even a notification is sent just after a few minutes after an application is being used, the attentional states or engagement/challenge levels can be detected successfully.

This study showed that GLMM fits with Bayesian approach may be preferable rather than individual random forest when there is not sufficient data available since GLMM does not require a high number of data points, and it also incorporates random effects itself. This may be a solution for a cold-start problem stated in [18]. GLMM also fits a population-level mean; hence, when a new user is added to the system, at first, population-level predictions could be used. Instead of fitting a general random forest model, and separate individual random forest models, GLMM may handle both.

CHAPTER 7

CONCLUSION

Ubiquitous technologies enable mobile developers and researchers to understand mobile users' context information appropriately. Several efforts have been made specifically for the inference of available moments of the users in order to send effective mobile notifications. It has been showed that responsiveness of the users to mobile notifications is affected by the environmental and personal factors. Researchers of previous studies focus on how those factors could be inferred from mobile or wearable sensors, and how effective models could be built upon the data obtained from those sensors. Recently, building effective personalized models has become popular in the domain of mobile computing and machine learning.

Surely, there are handicaps for building personalized models, especially at the very beginning of modelling phase. This is due to the limited number of data points in that phase. As the number of data points increases with the duration of the data collection period, more effective models can be generated upon the high quantity data collected. This situation is known as the "cold-start". Several solutions have been offered to that problem, such as using generalized models at the beginning of the data modelling.

In Chapter 4, a hybrid model has been proposed using Generalized Linear Mixed Model approach for the inference of rest break availability of office workers. The model, first, considers the time information regarding self-reported break availability responses. Kernel density estimations are computed based on the time and self-reported data. Then, density estimations and mobile sensor data namely location, physical activity, and ringer mode of the phone are used for predicting the break availability. In this phase, GLMM enable modelling at within- and between-subject level, hence, individualized predictions could be obtained. The model is compared with the well-known methods used in the literature, i.e., random forest models, both at individual-level and the population-level. The results show the efficiency of GLMM compared to them.

In Chapter 5, a framework including personal and social factors of office workers has been proposed and validated. In this study, engagement and challenge levels of office workers are considered as the target. Specifically, the response variations caused by individual characteristics of participants are explained with the framework. The factors included in the framework are musculoskeletal discomfort, awareness about rest breaks, personality traits and mobile application usage. As far as we know, this is the first of its kind to bring together those factors. Another contribution of the study is that engagement and challenge responses of the participants are investigated in terms of well-known responsiveness metrics including acquiescence, middle response style, entropy, and polarization. Significant relationships have been found and discussed.

Finally, the relation between *in-situ* engagement/challenge levels and the mobile application usage have been explored using repeated-measures correlation in Chapter 6. In addition, the in-situ engagement/challenge levels and attentional states have been modelled with GLMM. Similar to the first study, the comparison of GLMM with common methods has been provided. The efficiency of GLMM has been validated once again. The results show that mobile application usage metrics such as number of switches between mobile applications or the duration of mobile application usage are influential on predicting work engagement/challenge levels and attentional states of office workers.

The results of the studies included in the thesis have emphasized the importance of personalized data in a mobile system. The personal, environmental and social factors could enhance the effectiveness of such systems. The contributions of the thesis can be summarized as follows:

- Using a hybrid model including Generalized Linear Mixed Models, which address both within- and between-subjects' factors, is an efficient method for modelling the unbalanced and limited data. This approach can be employed at the "cold-start" period.
- The features calculated regarding user locations have a significant effect on the break availability of the participants. Hence, the features such as the duration spent in a location, the visit frequency of a location could be incorporated in the studies related to availability prediction.
- The factors such as musculoskeletal discomfort, the level of awareness, and subjective norm have not been explored before in terms of their effectiveness on the work engagement and challenge levels.
- User responses are investigated in terms of several metrics, e.g., acquiescence, disacquiescence, and extreme response style, which have not been considered in previous studies.
- The results of the study give insights regarding office environments, personality of office workers, and mobile phone usage during work hours. Effective mobile notification systems might be built considering those factors.
- The methods considering repeated-measures nature of the data set have given significant results without the violation of the statistical assumptions, so that more reliable results could be obtained.

7.1 Implications

As can be seen from the contributions listed above, researchers who would like to study on this topic in future, and mobile designers could benefit from the results of the study in terms of the factors that should be considered when developing an effective mobile notification system. The hybrid model proposed in Chapter 4 can be applied to a variety of office environments given its generality. In order to promote physical exercises or taking simple rest breaks in office environments, the context information should be considered. Mobile designs that remind office workers to take breaks based on the time, current location, ringer mode status, current activity might have a higher chance to be effective on the users.

The results presented in Chapter 5 provide insights related to personal and social factors effective on work engagement and challenge levels. First of all, as the previous studies demonstrated, office environment (e.g., office type, number of colleagues) gives clues about the work style of office workers. Distractions generated in an office environment, reactions to those distractions, and employees' attentional states caused by the distractions might differ with the office type. For these reasons, office-related factors should be considered before sending health-related interventions through mobile phones to knowledge workers. Besides, as the results of the study show, the degree of being influenced by co-workers (in terms of social norms) changes with the number of co-workers around.

Secondly, mobile notifications should comprise personal information based on the personality characteristics of users. Work engagement/challenge levels of employees could be predicted with the personality type so that suggesting more personalized time intervals for rest breaks might be one of the practical implications.

Finally, in Chapter 6, it has been once more shown that mobile application usage is a successful indicator for measuring work-related states such as "bored" or "focused" states. It may be quite useful to focus on both longer (e.g. up to one hour) and short amount of time when investigating application usage for the inference of work engagement, challenge, and attentional states. Further investigations could benefit from this finding.

7.2 Limitations and Future Work

Several points need a further investigation despite all the contributions listed above. First of all, the hybrid model proposed lacks accurate location tracking, because of that user locations are clustered based on Wi-Fi access points. Even though several threshold values are tested for location identification, the results could be partially affected because the calculations depend on the signals of Wi-Fi access points. In addition, the results could potentially be improved if the exact locations of the users could be obtained. Consequently, the location-dependent variables (e.g., time spent in location and location frequency) could also be affected from obtaining exact locations.

The features in the first study are limited because they are obtained from only mobile phone sensors to predict the availability for rest breaks. Similarly, in the third study, the main focus is only the mobile application usage variables for predicting attentional states and engagement/challenge levels. Computer activity or biometric sensors could be benefited to make more accurate predictions with a combination of mobile phone sensors. A further study could be employed with those sensors in addition to mobile phone sensors.

A more generalized model could be built and the results of the correlation analyses could be validated with a larger amount of data points as a future work. Although GLMM can handle limited number of data points, there is extremely low number of data points for some users, and that caused individual models not to learn those users' characteristics well. The main reason for not achieving high number of data points could be that responding to six ESM messages in each experiment day might be troublesome for those participants. In addition, because of the limited number of data points, the data set did not cover sufficient data points for different activity types (e.g., driving, walking); hence, activity had to be classified into two categories: "still" and "moving". With a larger amount of data, the effects of different activity types on break availability could be investigated separately. Besides, the number of data points in the categories of engagement/challenge levels was also unbalanced (e.g., 132 in category 1 of engagement, 52 in category 3 of engagement). Because of that, classifiers might not be able to learn that category as well as the other categories with a higher number of data. An increasing number of data points may lead to more balanced categories, or at least it might enable the classifier to learn categories better.

Model performances could be affected by the parameters used in the calculations. As mentioned above, selecting threshold value might be an important step because threshold affects the calculations of location parameters. Similarly, activity prediction for users without a significant sensor could also affect the results because there might be a chance to misclassify their activity. Those parameters might affect the model's sensitivity.

Another limitation of the study is the composition of the participants. The results of this study may not be representative of the general population due to the limitation of the convenience sampling, only office workers working in one city could be reached. The participants may have shared common personality characteristics already, hence, a more comprehensive study might be needed to validate the results.

Finally, data obtained from the questionnaires (e.g. MPPUS and the post-questionnaire) could not be analyzed since the number of participants who filled in those questionnaires is relatively low. Those measures could be obtained from a higher number of participants with a more extensive experiment as a future work, which enables exploring the relations with mobile phone addiction level, and the usability of the mobile sensing application.

REFERENCES

[1] J. Ho and S. S. Intille, "Using context-aware computing to reduce the perceived burden of interruptions from mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, pp. 909–918, 2005.

[2] G. Mark, S. T. Iqbal, M. Czerwinski, and P. Johns, "Bored mondays and focused afternoons: the rhythm of attention and online activity in the workplace," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3025–3034, 2014.

[3] D. M. Cutler, "Behavioral health interventions: what works and why?," in *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life*, pp. 643–674, Washington, D.C.: The National Academy Press, 2004.

[4] B. J. Fogg, "Persuasive technologies," *Communications of the ACM*, vol. 42, pp. 26–29, 5 1999.

[5] B. J. Fogg, "A behavior model for persuasive design," *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*, p. 40, 2009.

[6] H. Oinas-Kukkonen, "Behavior change support systems: a research model and agenda," in *Persuasive Technology*, pp. 4–14, Springer Berlin Heidelberg, 2010.

[7] H. Stern, V. Pammer, and S. Lindstaedt, "A preliminary study on interruptibility detection based on location and calendar information," in *Proceedings of the 3rd Workshop on Context-Systems Design, Evaluation and Optimisation*, 2011.

[8] G. N. Healy, D. W. Dunstan, J. Salmon, E. Cerin, J. E. Shaw, P. Z. Zimmet, and N. Owen, "Breaks in sedentary time: beneficial associations with metabolic risk," *Diabetes Care*, vol. 31, pp. 661–6, 4 2008.

[9] T. Y. Warren, V. Barry, S. P. Hooker, X. Sui, T. S. Church, and S. N. Blair, "Sedentary behaviors increase risk of cardiovascular disease mortality in men," *Medicine and Science in Sports and Exercise*, vol. 42, pp. 879–85, 5 2010.

[10] R. A. Henning, P. Jacques, G. V. Kissel, A. B. Sullivan, and S. M. Alteras-Webb, "Frequent short rest breaks from computer work: effects on productivity and well-being at two field sites," *Ergonomics*, vol. 40, pp. 78–91, 1 1997.

[11] L. McLean, M. Tingley, R. Scott, and J. Rickards, "Computer terminal work and the benefit of microbreaks," *Applied Ergonomics*, vol. 32, pp. 225–237, 6 2001.

[12] R. D. V. Barredo and K. Mahon, "The effects of exercise and rest breaks on musculoskeletal discomfort during computer tasks: an evidence-based perspective," *Journal of Physical Therapy Science*, vol. 19, no. 2, pp. 151–163, 2007.

[13] D. Morris, A. B. Brush, and B. R. Meyers, "SuperBreak: using interactivity to enhance ergonomic typing breaks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (New York, USA), pp. 1817–1826, ACM Press, 4 2008.

[14] Z. Podniece, S. Heuvel, and B. Blatter, *Work-related musculoskeletal disorders: prevention report*. European Agency for Safety and Health at Work, 2008.

[15] C. Leah, "Exercises to reduce musculoskeletal discomfort for people doing a range of static and repetitive work," tech. rep., Health and Safety Laboratory, 2011.

[16] M. Friedrich, G. Gittler, Y. Halberstadt, T. Cermak, and I. Heiller, "Combined exercise and motivation program: effect on the compliance and level of disability of patients with chronic low back pain: a randomized controlled trial," *Archives of Physical Medicine and Rehabilitation*, vol. 79, pp. 475–487, 5 1998.

[17] T. Galinsky, N. Swanson, S. Sauter, R. Dunkin, J. Hurrell, and L. Schleifer, "Supplementary breaks and stretching exercises for data entry operators: a follow-up field study," *American Journal of Industrial Medicine*, vol. 50, pp. 519–27, 7 2007.

[18] M. Züger, S. C. Müller, A. N. Meyer, and T. Fritz, "Sensing interruptibility in the office: a field study on the use of biometric and computer interaction sensors," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 591, 2018.

[19] A. Zenonos, A. Khan, G. Kalogridis, S. Vatsikas, T. Lewis, and M. Sooriyabandara, "HealthyOffice: mood recognition at work using smartphones and wearable sensors," in *2016 IEEE International Conference on Pervasive Computing and Communication Workshops*, pp. 1–6, IEEE, 3 2016.

[20] B. Brown, M. McGregor, and D. McMillan, "100 days of iPhone use," in *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices and Services*, (New York, New York, USA), pp. 223–232, ACM Press, 2014.

[21] A. Matic, M. Pielot, and N. Oliver, "Boredom-computer interaction: boredom proneness and the use of smartphone," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 837–841, 2015.

[22] M. Pielot, T. Dingler, J. S. Pedro, and N. Oliver, "When attention is not scarce - detecting boredom from mobile phone usage," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 825–836, ACM, 2015.

[23] R. Bixler and S. D'Mello, "Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits," in *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, (New York, USA), pp. 225–234, ACM Press, 2013.

[24] D. L. Paulhus, "Measurement and control of response bias," in *Measures of Personality and Social Psychological Attitudes*, pp. 17–59, San Diego: Elsevier, 1991.

[25] D. C. Turk and T. E. Rudy, "Neglected topics in the treatment of chronic pain patients — relapse, noncompliance, and adherence enhancement," *Pain*, vol. 44, no. 1, pp. 5–28, 1991.

[26] Y. Luo, B. Lee, D. Y. Wohn, A. L. Rebar, D. E. Conroy, and E. K. Choe, "Time for break: understanding information workers' sedentary behavior through a break prompting system," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2018.

[27] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, 1991.

[28] I. Ajzen and M. Fishbein, *Understanding Attitudes and Predicting Social Behavior*. Prentice-Hall, 1980.

[29] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley, MA, 1975.

[30] S. Van Dantzig, G. Geleijnse, and A. T. Van Halteren, "Toward a persuasive mobile application to reduce sedentary behavior," *Personal and Ubiquitous Computing*, vol. 17, no. 6, pp. 1237–1246, 2013.

[31] E. S. George, G. S. Kolt, R. R. Rosenkranz, and J. M. Guagliano, "Physical activity and sedentary time: male perceptions in a university work environment," *American Journal of Men's Health*, vol. 8, no. 2, pp. 148–158, 2014.

[32] R. L. Morrison and K. A. Macky, "The demands and resources arising from shared office spaces," *Applied Ergonomics*, vol. 60, pp. 103–115, 2017.

[33] S. L. Mullane, M. J. Toledo, S. A. Rydell, L. H. Feltes, B. Vuong, N. C. Crespo, M. A. Pereira, and M. P. Buman, "Social ecological correlates of workplace sedentary behavior," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 14, no. 1, p. 117, 2017.

[34] A. Visuri, N. Van Berkel, C. Luo, J. Goncalves, D. Ferreira, and V. Kostakos, "Predicting interruptibility for manual data collection: a cluster-based user model," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, p. 12, ACM, 2017.

[35] M. Pielot, L. Baltrunas, and N. Oliver, "Boredom-triggered proactive recommendations," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pp. 1106–1110, 2015.

[36] A. Mehrotra, M. Musolesi, R. Hendley, and V. Pejovic, "Designing content-driven intelligent notification mechanisms for mobile applications," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 813–824, 2015.

[37] J. Z. Bakdash and L. R. Marusich, "Repeated measures correlation," *Frontiers in Psychology*, vol. 8, no. MAR, pp. 1–13, 2017.

[38] L. Coraggio, *Deleterious Effects of Intermittent Interruptions on the Task Performance of Knowledge Workers: A Laboratory Investigation*. PhD thesis, University of Arizona, 1990.

[39] C. Speier, J. S. Valacich, and I. Vessey, "The effects of task interruption and information presentation on individual decision making," in *Proceedings of the 18th International Conference on Information Systems*, pp. 21–36, Association for Information Systems, 12 1997.

[40] S. A. Grandhi and Q. Jones, "Conceptualizing interpersonal interruption management: a theoretical framework and research program," in *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences*, pp. 1–10, 2009.

[41] K. Glanz and D. B. Bishop, "The role of behavioral science theory in development and implementation of public health interventions," *Annual Review of Public Health*, vol. 31, pp. 399–418, 3 2010.

[42] G. M. Hochbaum, *Public Participation in Medical Screening Programs: A Socio-Psychological Study*. US Department of Health, Education, and Welfare, 1958.

[43] V. L. Champion and C. S. Skinner, "The health belief model," in *Health Behavior and Health Education: Theory, Research, and Practice*, pp. 45–65, San Francisco: Jossey-Bass, 2008.

[44] J. O. Prochaska and C. C. DiClemente, *The Transtheoretical Approach: Crossing Traditional Boundaries of Therapy*. Homewood, IL: Dorsey Press, 1984.

[45] A. Dijkstra, H. De Vries, and J. Roijackers, "Targeting smokers with low readiness to change with tailored and non-tailored self-help materials," *Preventive Medicine*, vol. 28, pp. 203–211, 1999.

[46] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory*. Eaglewood Cliffs, New York: Prentice-Hall, 1986.

[47] A. L. McAlister, C. L. Perry, and G. S. Parcel, "How individuals, environments and health behaviors interact: social cognitive theory," in *Health Behavior and Health Education: Theory, Research, and Practice*, pp. 169–188, San Francisco: Jossey-Bass, 2008.

[48] K. R. McLeroy, D. Bibeau, A. Steckler, and K. Glanz, "An ecological perspective on health promotion programs," *Health Education Quarterly*, vol. 15, pp. 351–377, 12 1988.

[49] H. Oinas-Kukkonen and M. Harjumaa, "Towards deeper understanding of persuasion in software and information systems," in *First International Conference on Advances in Computer-Human Interaction*, pp. 200–205, IEEE, 2008.

[50] W. J. McGuire, "Persuasion," in *Communication, Language, and Meaning Psychological Perspectives*, pp. 242–255, New York: Basic Books, 1973.

[51] C. Fraser, B. Burchell, and D. Hay, *Introducing Social Psychology*. Cambridge: Polity, 2001.

[52] R. Petty and J. T. Cacioppo, *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*. New York: Springer, 1986.

[53] R. Cialdini, *Influence, Science and Practice*. Allyn and Bacon, 2001.

[54] H. W. Simons, J. Morreale, and B. Gronbeck, *Persuasion in Society*. Thousand Oaks: SAGE Publications, 2001.

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

[56] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive systems design: key issues, process model, and system features," *Communications of the Association for Information Systems*, vol. 24, pp. 485–500, 2009.

[57] H. Oinas-Kukkonen, "A foundation for the study of behavior change support systems," *Personal and Ubiquitous Computing*, vol. 17, pp. 1223–1235, 7 2012.

[58] V. Strecher, "Internet methods for delivering behavioral and health-related interventions (eHealth)," *Annual Review of Clinical Psychology*, vol. 3, pp. 53–76, 1 2007.

[59] M. Oduor, T. Alahäivälä, and H. Oinas-Kukkonen, "Persuasive software design patterns for social influence," *Personal and Ubiquitous Computing*, vol. 18, pp. 1689–1704, 7 2014.

[60] M. Kaptein, P. Markopoulos, B. De Ruyter, and E. Aarts, "Can you be persuaded? Individual differences in susceptibility to persuasion," in *Human-Computer Interaction–INTERACT 2009*, pp. 115–118, 2009.

[61] M. Kaptein, J. Lacroix, and P. Saini, "Individual differences in persuadability in the health promotion domain," in *Persuasive Technology*, pp. 94–105, Springer, 2010.

[62] M. Kaptein, P. Markopoulos, B. de Ruyter, and E. Aarts, "Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles," *International Journal of Human-Computer Studies*, vol. 77, pp. 38–51, 2015.

[63] M. Kaptein and A. Van Halteren, "Adaptive persuasive messaging to increase service retention: using persuasion profiles to increase the effectiveness of email reminders," *Personal and Ubiquitous Computing*, vol. 17, no. 6, pp. 1173–1185, 2013.

[64] V. Pejovic and M. Musolesi, "Anticipatory mobile computing for behaviour change interventions," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, (New York, USA), pp. 1025–1034, ACM Press, 9 2014.

[65] N. Lathia, V. Pejovic, K. Rachuri, C. Mascolo, M. Musolesi, and P. J. Rentfrow, "Smartphones for large-scale behaviour change interventions," *IEEE Pervasive Computing*, pp. 1–11, 2014.

[66] R. Rosen, *Anticipatory Systems*. Oxford, UK: Pergamon Press, 1985.

[67] A. Madan, M. Cebrian, D. Lazer, and A. Pentland, "Social sensing for epidemiological behavior change," in *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, (New York, New York, USA), p. 291, ACM Press, 9 2010.

[68] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, (New York, USA), pp. 3–14, ACM Press, 9 2014.

[69] K. Farrahi, R. Emonet, and M. Cebrian, "Predicting a community's flu dynamics with mobile phone data," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, (New York, USA), pp. 1214–1221, ACM Press, 2 2015.

[70] M. Csikszentmihalyi and R. Lardon, "The experience sampling method," *New Directions for Methodology of Social and Behavioral Science*, vol. 15, pp. 41–56, 1983.

[71] V. Pejovic, N. Lathia, C. Mascolo, and M. Musolesi, "Mobile-based experience sampling for behaviour research," in *Emotions and Personality in Personalized Services*, pp. 141–161, Springer, Cham, 8 2016.

[72] Y.-J. Chang and J. C. Tang, "Investigating mobile users' ringer mode usage and attentiveness and responsiveness to communication," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 6–15, 2015.

[73] M. Pielot, B. Cardoso, K. Katevas, J. Serrà, A. Matic, and N. Oliver, "Beyond interruptibility: predicting opportune moments to engage mobile phone users," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–25, 2017.

[74] A. Mathur, N. D. Lane, and F. Kawsar, "Engagement-aware computing: modelling user engagement from mobile contexts," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 622–633, 2016.

[75] A. Mehrotra, V. Pejovic, J. Vermeulen, R. Hendley, and M. Musolesi, "My phone and me: understanding people's receptivity to mobile notifications," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1021–1032, 2016.

[76] F. Yuan, X. Gao, and J. Lindqvist, "How busy are you? Predicting the interruptibility intensity of mobile users," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5346–5360, ACM, 2017.

[77] J. He, D. Bartram, I. Inceoglu, and F. J. R. van de Vijver, "Response styles and personality traits: a multilevel analysis," *Journal of Cross-Cultural Psychology*, vol. 45, no. 7, pp. 1028–1045, 2014.

[78] A.-W. Harzing, "Response styles in cross-national survey research: A 26-country study," *International Journal of Cross Cultural Management*, vol. 6, no. 2, pp. 243–266, 2006.

[79] A. Couch and K. Keniston, "Yeasayers and naysayers: Agreeing response set as a personality variable," *The Journal of Abnormal and Social Psychology*, vol. 60, no. 2, pp. 151–174, 1960.

[80] M. V. Hibbing, M. Cawvey, R. Deol, A. J. Bloeser, and J. J. Mondak, "The relationship between personality and response patterns on public opinion surveys: The Big Five, extreme response style, and acquiescence response style," *International Journal of Public Opinion Research*, vol. 31, no. 1, pp. 161–177, 2019.

[81] J. He and F. J. van de Vijver, "A general response style factor: Evidence from a multi-ethnic study in the Netherlands," *Personality and Individual Differences*, vol. 55, no. 7, pp. 794–800, 2013.

[82] J. R. Hurley, "Timidity as a response style to psychological questionnaires," *The Journal of Psychology*, vol. 132, no. 2, pp. 201–210, 1998.

[83] E. J. Austin, I. J. Deary, and V. Egan, "Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items," *Personality and Individual Differences*, vol. 40, no. 6, pp. 1235–1245, 2006.

[84] R. Whittaker, H. McRobbie, C. Bullen, R. Borland, A. Rodgers, and Y. Gu, "Mobile phone-based interventions for smoking cessation," *Cochrane Database of Systematic Reviews*, no. 11, 2012.

[85] S.-I. Kim and H.-S. Kim, "Effectiveness of mobile and internet intervention in patients with obese type 2 diabetes.," *International Journal of Medical Informatics*, vol. 77, pp. 399–404, 6 2008.

[86] S. Consolvo, R. Libby, I. Smith, J. A. Landay, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, and L. LeGrand, “Activity sensing in the wild,” in *Proceeding of the 26th Annual CHI Conference on Human Factors in Computing Systems*, (New York, USA), p. 1797, ACM Press, 4 2008.

[87] V. Rajanna, D. J. Behera, D. Goldberg, and T. Hammond, “Step up life: a context aware health assistant,” in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health*, pp. 21–30, 2014.

[88] P. Klasnja and W. Pratt, “Healthcare in the pocket: mapping the space of mobile-phone health interventions,” *Journal of Biomedical Informatics*, vol. 45, no. 1, pp. 184–198, 2012.

[89] C. Free, G. Phillips, L. Galli, L. Watson, L. Felix, P. Edwards, V. Patel, and A. Haines, “The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review,” *PLoS Medicine*, vol. 10, no. 1, 2013.

[90] B. S. Fjeldsoe, Y. D. Miller, and A. L. Marshall, “MobileMums: a randomized controlled trial of an SMS-based physical activity intervention,” *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, vol. 39, pp. 101–11, 5 2010.

[91] L. Pina, E. Ramirez, and W. Griswold, “Fitbit+: a behavior-based intervention system to reduce sedentary behavior,” in *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 175–178, 2012.

[92] L. E. Burke, M. B. Conroy, S. M. Sereika, O. U. Elci, M. a. Styn, S. D. Acharya, M. a. Sevick, L. J. Ewing, and K. Glanz, “The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial.,” *Obesity*, vol. 19, no. 2, pp. 338–344, 2011.

[93] H. op den Akker, M. Cabrita, R. op den Akker, V. M. Jones, and H. J. Hermens, “Tailored motivational message generation: a model and practical framework for real-time physical activity coaching,” *Journal of Biomedical Informatics*, vol. 55, pp. 104–115, 2015.

[94] M. Tabak, H. Op den Akker, and H. Hermens, “Motivational cues as real-time feedback for changing daily activity behavior of patients with COPD,” *Patient Education and Counseling*, vol. 94, no. 3, pp. 372–378, 2014.

[95] E. Schneider, X. Irastorza, and S. Copsey, “OSH in figures: Work-related musculoskeletal disorders in the EU - Facts and figures,” tech. rep., European Agency for Safety and Health at Work, 2010.

[96] R. Williams and M. Westmorland, “Occupational cumulative trauma disorders of the upper extremity,” *American Journal of Occupational Therapy*, vol. 48, pp. 411–420, 5 1994.

[97] K. T. Palmer, “Use of keyboards and symptoms in the neck and arm: evidence from a national survey,” *Occupational Medicine*, vol. 51, pp. 392–395, 9 2001.

[98] M. Monsey, I. Ioffe, A. Beatini, B. Lukey, A. Santiago, and A. B. James, “Increasing compliance with stretch breaks in computer users through reminder software,” *Work (Reading, Mass.)*, vol. 21, pp. 107–11, 1 2003.

[99] L. Trujillo and X. Zeng, "Data entry workers perceptions and satisfaction response to the "stop and stretch" software program," *Work (Reading, Mass.)*, vol. 27, pp. 111–21, 1 2006.

[100] D. Berque, J. Burgess, A. Billingsley, S. Johnson, T. L. Bonebright, and B. Wethington, "Design and evaluation of persuasive technology to encourage healthier typing behaviors," in *Proceedings of the 6th International Conference on Persuasive Technology Persuasive Technology and Design: Enhancing Sustainability and Health*, (New York, USA), pp. 1–10, ACM Press, 6 2011.

[101] D. Cooley and S. Pedersen, "A pilot study of increasing nonpurposeful movement breaks at work as a means of reducing prolonged sitting," *Journal of Environmental and Public Health*, vol. 2013, 2013.

[102] S.-C. Wang, C. Jiang, and J.-Y. Chern, "Promoting healthy computer use: Timing-informed computer health animations for prolonged sitting computer users," *Behaviour & Information Technology*, vol. 33, no. 3, pp. 295–301, 2014.

[103] D. A. Epstein, D. Avrahami, and J. T. Biehl, "Taking 5: Work-breaks, productivity, and opportunities for personal informatics for knowledge workers," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 673–684, 2016.

[104] R. M. Mabry, Z. Q. Al-Busaidi, M. M. Reeves, N. Owen, and E. G. Eakin, "Addressing physical inactivity in Omani adults: perceptions of public health managers," *Public Health Nutrition*, vol. 17, no. 3, pp. 674–681, 2013.

[105] B. Wallmann-Sperlich, J. Bucksch, S. Schneider, and I. Froboese, "Socio-demographic, behavioural and cognitive correlates of work-related sitting time in German men and women," *BMC Public Health*, vol. 14, no. 1, p. 1259, 2014.

[106] W. B. Schaufeli, A. B. Bakker, and M. Salanova, "The measurement of work engagement with a short questionnaire," *Educational and Psychological Measurement*, vol. 66, pp. 701–716, 8 2006.

[107] O. Fenichel, "On the psychology of boredom," in *Organization and Pathology of Thought: Selected Sources*, pp. 349–361, New York: Columbia University Press, 1951.

[108] C. D. Fisherl, "Boredom at work: A neglected concept," *Human Relations*, vol. 46, no. 3, pp. 395–417, 1993.

[109] T. Goetz, A. C. Frenzel, N. C. Hall, U. E. Nett, R. Pekrun, and A. A. Lipnevich, "Types of boredom: An experience sampling approach," *Motivation and Emotion*, vol. 38, pp. 401–419, 6 2014.

[110] R. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[111] G. Mark, S. Iqbal, M. Czerwinski, and P. Johns, "Capturing the mood: Facebook and face-to-face encounters in the workplace," in *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & social Computing*, pp. 1082–1094, 2014.

[112] A. B. Bakker, E. Demerouti, and L. L. ten Brummelhuis, "Work engagement, performance, and active learning: The role of conscientiousness," *Journal of Vocational Behavior*, vol. 80, no. 2, pp. 555–564, 2012.

[113] F.-Y. Liao, L.-Q. Yang, M. Wang, D. Drown, and J. Shi, “Team–member exchange and work engagement: Does personality make a difference?,” *Journal of Business and Psychology*, vol. 28, pp. 63–77, 3 2013.

[114] G. Zecca, C. Györkös, J. Becker, K. Massoudi, G. de Bruin, and J. Rossier, “Validation of the French Utrecht Work Engagement Scale and its relationship with personality traits and impulsivity,” *European Review of Applied Psychology*, vol. 65, pp. 19–28, 1 2015.

[115] G. Mark, S. T. Iqbal, M. Czerwinski, and P. Johns, “Focused, aroused, but so distractible: A temporal perspective on multitasking and communications,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 903–916, ACM, 2015.

[116] D. Nduhura and M. Prieler, “When I chat online, I feel relaxed and work better: Exploring the use of social media in the public sector workplace in Rwanda,” *Telecommunications Policy*, vol. 41, no. 7-8, pp. 708–716, 2017.

[117] S. Mirjafari, K. Masaba, T. Grover, W. Wang, P. Audia, A. T. Campbell, N. V. Chawla, V. D. Swain, M. De Choudhury, A. K. Dey, S. K. D’Mello, G. Gao, J. M. Gregg, K. Jagannath, K. Jiang, S. Lin, Q. Liu, G. Mark, G. J. Martinez, S. M. Mattingly, E. Moskal, R. Mulukutla, S. Nepal, K. Nies, M. D. Reddy, P. Robles-Granda, K. Saha, A. Sirigiri, and A. Striegel, “Differentiating higher and lower job performers in the workplace using mobile sensing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, pp. 1–24, 6 2019.

[118] W. Choi, S. Park, D. Kim, Y.-k. Lim, and U. Lee, “Multi-stage receptivity model for mobile just-in-time health intervention,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, pp. 1–26, 2019.

[119] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, “MoodScope: Building a mood sensor from smartphone usage patterns,” in *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, (New York, USA), pp. 389–402, ACM Press, 2013.

[120] W. C. Taylor, R. Shegog, V. Chen, D. M. Rempel, M. P. Baun, C. L. Bush, T. Green, and N. Hare-Everline, “The Booster Break program: Description and feasibility test of a worksite physical activity daily practice,” *Work*, vol. 37, no. 4, pp. 433–443, 2010.

[121] V. Venkatesh and F. D. Davis, “A theoretical extension of the technology acceptance model: Four longitudinal field studies,” *Management Science*, vol. 46, no. 2, pp. 186–204, 2000.

[122] H. Mohd Mohadis and N. Mohamad Ali, “Smartphone application for physical activity enhancement at workplace: Would office workers actually use it?,” in *2018 International Conference on Information and Communication Technology for the Muslim World*, pp. 144–149, IEEE, 7 2018.

[123] A. Seddigh, E. Berntson, L. G. Platts, and H. Westerlund, “Does personality have a different impact on self-rated distraction, job satisfaction, and job performance in different office types?,” *PLOS ONE*, vol. 11, no. 5, 2016.

[124] S. Kucukozer Cavdar and T. Taskaya Temizel, “A framework for investigating the factors that affect knowledge workers to take rest breaks in work hours,” in *8th International Conference on Occupational Safety & Health*, (Istanbul), 2016.

[125] E. Akkurt, *UBDroid: A Tool for Monitoring Smartphone Application Usage for User Behavior Analysis*. PhD thesis, Middle East Technical University, 2015.

[126] S. Kucukozer Cavdar, E. Akkurt, D. Cavdar, A. Temizel, and T. Taskaya Temizel, "UBDroid: Kullanıcı davranış analizi için akıllı telefon uygulamaları kullanım izleme aracı," in *10.Ulusal Yazılım Mühendisliği Sempozyumu (UYMS)*, (Canakkale, Turkey), 2016.

[127] Google, "Google Activity Recognition API," 2019.

[128] P. Yu, H. Li, and M.-P. Gagnon, "Health IT acceptance factors in long-term care facilities: A cross-sectional survey," *International Journal of Medical Informatics*, vol. 78, no. 4, pp. 219–229, 2009.

[129] M.-C. Hung and W.-Y. Jen, "The adoption of mobile health management services: An empirical study," *Journal of Medical Systems*, vol. 36, pp. 1381–1388, 6 2012.

[130] B. Kijsanayotin, S. Pannarunothai, and S. M. Speedie, "Factors influencing health information technology adoption in Thailand's community health centers: Applying the UTAUT model," *International Journal of Medical Informatics*, vol. 78, no. 6, pp. 404–416, 2009.

[131] T. Gencoz and O. Oncul, "The journal of general examination of personality characteristics in a Turkish sample: Development of basic personality traits inventory," *The Journal of General Psychology*, vol. 139, no. 3, pp. 194–216, 2012.

[132] O. P. John and S. Srivastava, "The Big Five Trait taxonomy: History, measurement, and theoretical perspectives," in *Handbook of Personality: Theory and Research* (L. A. Pervin and O. P. John, eds.), pp. 102–138, New York: Guilford Press, 1999.

[133] R. R. McCrae and P. T. Costa, "Validation of the five-factor model of personality across instruments and observers," *Journal of Personality and Social Psychology*, vol. 52, no. 1, pp. 81–90, 1987.

[134] C. Durrett and T. J. Trull, "An evaluation of evaluative personality terms: A comparison of the Big Seven and Five-factor model in predicting psychopathology," *Psychological Assessment*, vol. 17, no. 3, pp. 359–368, 2005.

[135] A. Bianchi and J. G. Phillips, "Psychological predictors of problem mobile phone use," *Cyberpsychology & Behavior*, vol. 8, no. 1, pp. 39–51, 2005.

[136] T. Koivumäki, A. Ristola, and M. Kesti, "The perceptions towards mobile services: An empirical analysis of the role of use facilitators," *Personal and Ubiquitous Computing*, vol. 12, no. 1, pp. 67–75, 2008.

[137] G. H. Ter Hofte, "Xensible interruptions from your mobile phone," in *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services*, pp. 178–181, 2007.

[138] J. E. Fischer, N. Yee, V. Bellotti, N. Good, S. Benford, and C. Greenhalgh, "Effects of content and time of delivery on receptivity to mobile interruptions," in *Proceedings of the 12th International Conference on Human-computer Interaction with Mobile Devices and Services*, pp. 103–112, 2010.

[139] V. Pejovic and M. Musolesi, "InterruptMe: designing intelligent prompting mechanisms for pervasive applications," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 897–908, ACM, 2014.

[140] B. Poppinga, W. Heuten, and S. Boll, "Sensor-based identification of opportune moments for triggering notifications," *IEEE Pervasive Computing*, vol. 13, no. 1, pp. 22–29, 2014.

[141] M. Pielot, K. Church, and R. de Oliveira, "An in-situ study of mobile phone notifications," in *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices and Services*, pp. 233–242, 2014.

[142] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt, "Large-scale assessment of mobile notifications," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3055–3064, 2014.

[143] H. Oh, L. Jalali, and R. Jain, "An intelligent notification system using context from real-time personal activity monitoring," in *IEEE International Conference on Multimedia and Expo*, pp. 1–6, 2015.

[144] A. Mehrotra, R. Hendley, and M. Musolesi, "PrefMiner: Mining user's preferences for intelligent mobile notification management," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*, pp. 1223–1234, ACM, 2016.

[145] J. Hadfield, "MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package," *Journal of Statistical Software*, vol. 33, no. 2, pp. 1–23, 2010.

[146] C. E. Mcculloch, "Generalized linear mixed models," in *NSF-CBMS Regional Conference Series in Probability and Statistics*, Institute of Mathematical Statistics, 2003.

[147] J. D. Hadfield and S. Nakagawa, "General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters," *Journal of Evolutionary Biology*, vol. 23, no. 3, pp. 494–508, 2010.

[148] H. T. Ventura, F. F. e. Silva, L. Varona, E. A. P. d. Figueiredo, E. V. Costa, L. P. d. Silva, R. Ventura, and P. S. Lopes, "Comparing multi-trait Poisson and Gaussian Bayesian models for genetic evaluation of litter traits in pigs," *Livestock Science*, vol. 176, pp. 47–53, 2015.

[149] B. M. Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J. S. S. White, "Generalized linear mixed models: A practical guide for ecology and evolution," *Trends in Ecology and Evolution*, vol. 24, no. 3, pp. 127–135, 2009.

[150] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 583–639, 2002.

[151] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, FL, USA: Chapman & Hall, 2nd ed. ed., 2004.

[152] E. Aarts, M. Verhage, J. V. Veenvliet, C. V. Dolan, and S. Van Der Sluis, "A solution to dependency: Using multilevel analysis to accommodate nested data," *Nature Neuroscience*, vol. 17, no. 4, pp. 491–496, 2014.

[153] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge, 2002.

[154] J. M. Bland and D. G. Altman, “Calculating correlation coefficients with repeated observations: Part 1 - correlation within subjects,” *BMJ*, vol. 310, p. 446, 1995.

[155] J. M. Bland and D. G. Altman, “Calculating correlation coefficients with repeated observations: Part 2 - correlation between subjects,” *BMJ*, vol. 310, p. 633, 1995.

[156] Y. E. Ustev, O. Durmaz Incel, and C. Ersoy, “User, device and orientation independent human activity recognition on mobile phones,” in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct*, (New York, USA), pp. 1427–1436, ACM Press, 2013.

[157] M. Kose, O. D. Incel, and C. Ersoy, “Online human activity recognition on smart phones,” in *Proceedings of the 2nd International Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data*, pp. 11–15, 2012.

[158] H. Wang, A. Elgohary, and R. R. Choudhury, “No need to war-drive: Unsupervised indoor localization,” in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys ’12)*, pp. 197–210, 2012.

[159] N. Banovic, C. Brant, J. Mankoff, and A. Dey, “ProactiveTasks: The short of mobile device use sessions,” in *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices and Services*, (New York, USA), pp. 243–252, ACM Press, 2014.

[160] N. van Berkel, C. Luo, T. Anagnostopoulos, D. Ferreira, J. Goncalves, S. Hosio, and V. Kostakos, “A systematic assessment of smartphone usage gaps,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, (New York, USA), pp. 4711–4721, ACM Press, 2016.

[161] D. Ferreira, J. Goncalves, V. Kostakos, L. Barkhuus, and A. K. Dey, “Contextual experience sampling of mobile application micro-usage,” in *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, (New York, USA), pp. 91–100, ACM Press, 2014.

[162] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” *International Joint Conference on Artificial Intelligence*, vol. 14, no. 2, pp. 1137–1145, 1995.

[163] M. C. Jones, J. S. Marron, and S. J. Sheather, “A brief survey of bandwidth selection for density estimation,” *Journal of the American Statistical Association*, vol. 91, p. 401, 3 1996.

[164] J. D. Hadfield, “MCMCglmm: Markov chain Monte Carlo methods for generalised linear mixed models,” tech. rep., 2010.

[165] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[166] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[167] M. Friedman, “The use of ranks to avoid the assumption of normality implicit in the analysis of variance,” *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.

[168] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics*, pp. 196–202, New York, USA: Springer, 1992.

[169] L. D. Turner, S. M. Allen, and R. M. Whitaker, "Reachable but not receptive: Enhancing smartphone interruptibility prediction by modelling the extent of user engagement with notifications," *Pervasive and Mobile Computing*, vol. 40, pp. 480–494, 2017.

[170] A. Exler, M. Braith, A. Schankin, and M. Beigl, "Preliminary investigations about interruptibility of smartphone users at specific place types," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct (UbiComp '16)*, pp. 1590–1595, 2016.

[171] C. Park, J. Lim, J. Kim, S.-J. Lee, and D. Lee, ""Don't bother me. I'm socializing!": a breakpoint-based smartphone notification system," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pp. 541–554, 2017.

[172] T. Okoshi, H. Nozaki, J. Nakazawa, H. Tokuda, J. Ramos, and A. K. Dey, "Towards attention-aware adaptive notification on smart phones," *Pervasive and Mobile Computing*, vol. 26, pp. 17–34, 2016.

[173] J. E. Fischer, C. Greenhalgh, and S. Benford, "Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*, pp. 181–190, 2011.

[174] D. Ruedin, "An introduction to the R package agrmt," tech. rep., 2016.

[175] P. Van Kenhove, K. Wijnen, and K. De Wulf, "The influence of topic involvement on mail-survey response behavior," *Psychology and Marketing*, vol. 19, pp. 293–301, 3 2002.

[176] S. R. Porter and M. E. Whitcomb, "Non-response in student surveys: The role of demographics, engagement and personality," *Research in Higher Education*, vol. 46, pp. 127–152, 3 2005.

[177] A. Aviv, J. Zelenski, L. Rallo, and R. Larsen, "Who comes when: Personality differences in early and later participation in a university subject pool," *Personality and Individual Differences*, vol. 33, pp. 487–496, 8 2002.

[178] C. D. Stevens and R. A. Ash, "The conscientiousness of students in subject pools: Implications for "laboratory" research," *Journal of Research in Personality*, vol. 35, pp. 91–97, 3 2001.

[179] C. Montag, K. Błaszkiewicz, R. Sariyska, B. Lachmann, I. Andone, B. Trendafilov, M. Eibes, and A. Markowetz, "Smartphone usage in the 21st century: who is active on WhatsApp?," *BMC Research Notes*, vol. 8, p. 331, 12 2015.

[180] J. Smith and N. Dulay, "RingLearn: long-term mitigation of disruptive smartphone interruptions," in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, pp. 27–35, 2014.

[181] S. Patil, R. Hoyle, R. Schlegel, A. Kapadia, and A. J. Lee, "Interrupt now or inform later?: comparing immediate and delayed privacy feedback," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1415–1418, ACM, 2015.

[182] M. Choy, D. Kim, J.-G. Lee, H. Kim, and H. Motoda, “Looking back on the current day: Interruptibility prediction using daily behavioral features,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp ’16*, pp. 1004–1015, 2016.

[183] D. Weber, A. Voit, P. Kratzer, and N. Henze, “In-situ investigation of notifications in multi-device environments,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp ’16*, pp. 1259–1264, ACM, 2016.

[184] M. Obuchi, W. Sasaki, T. Okoshi, J. Nakazawa, and H. Tokuda, “Investigating interruptibility at activity breakpoints using smartphone activity recognition API,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 1602–1607, 2016.

[185] K. Komuro, Y. Fujimoto, and K. Fujita, “Relationship between worker interruptibility and work transitions detected by smartphone,” in *Human-Computer Interaction. Interaction Contexts. HCI 2017. Lecture Notes in Computer Science*, pp. 687–699, Springer, Cham, 2017.

[186] A. Morandi and R. Serafin, “A personalized motivation strategy for physical activity promotion in diabetic subjects,” in *Proceedings of the 2nd Workshop on Personalisation for E-Health*, pp. 51–55, 2007.

APPENDIX A

TABLES FOR RELATED WORK CHAPTER

A.1 Summary of the Interruption Studies

The table starts on next page.

Study	Objective/ Hypothesis	No of participants Data collection period	Scales used	Method	Significant features	Result
[1]	Users are more receptive to interruption at physical activity transitions than at random times	25 users One day	5-level Likert scale (Extremely receptive to Not at all receptive) was used in order to determine the receptivity of users	T-test conducted between two groups. One group received messages randomly, while the other received messages based on their activity information.	Activity of user is important on determining opportune moments to deliver interruption.	Participants were more receptive to interruptions during activity transitions than those delivered at a random time.
[137]	Investigating which context information predicts a user's availability for a phone call	10 users Seven days	A questionnaire sent by a mobile application.	Naïve Bayes	Social relation, being in a conversation, being in a face to face conversation, being at home	Opportune moments were predicted with 63.9% accuracy.
[138]	The receptivity is influenced by notification content and timing.	11 users 10 days	Participants rated their receptivity. The scale was not specified in detail.	ANOVA	Content	The reaction of participants for good content was significantly higher than for bad content.
[173]	Investigating whether opportune moments at the end of a mobile interaction (making voice call or receiving SMS).	20 users Two weeks	5-level Likert scale (Very burdensome to Not burdensome at all) was used in order to determine perceived workload of mobile tasks.	Chi-square	Time	Acceptance and response time are significantly higher for random times than opportune times.

Continued on next page.

Study	Objective/ Hypothesis	No of participants Data collection period	Scales used	Method	Significant features	Result
[141]	Investigating which application types' messages users respond to, in which phone mode (vibration, silent, or ringer) and in how many minutes they respond	15 users One week	A mobile app to record response time of users. A qualitative measurement called daily diary in order to keep subjective feedback	Mann-Whitney Test	Application type Ringer mode	Participants viewed messages from apps with messenger type significantly faster than e-mails, and they viewed notifications significantly faster in vibration mode than in silent and ringer mode.
[142]	Investigating message click time speed vs. application type	40191 users Six months	5-level Likert scale (Very important to Not important at all) was used in order to determine the importance of notifications to users.	ANOVA	Application category, importance of application category	System category had the shortest click time followed by messenger category. Reader/news category had the longest click time. Users rated notifications from messenger category as the most important ones.
[180]	Aims to predict opportune moments	10 users Two months	No scale was used. Only data was collected using a mobile phone application.	SVM, k-NN, Naïve Bayes	Day, time, contact number, SSID, GPS, cell tower ID	78.5% (SVM average), 71.2% (k-NN average), 71.7% (NB average)

Continued on next page.

Study	Objective/ Hypothesis	No of participants Data collection period	Scales used	Method	Significant features	Result
[140]	Aiming to predict opportune moments	79 users Five months	Two 5-level Likert scales (Strongly agree to Strongly disagree) were used in order to determine the mood of users and obtrusiveness of interruptions	Tree-based C4.5 classifier	Position, proximity, time, location provider	Opportune moments were classified with a 77.85% accuracy.
[181]	Investigating whether actionable, immediate or delayed notification types are more effective in terms of sharing personal information	35 users 15 days	5-level Likert scale (Uncomfortable to Comfortable) in order to collect users' feedback about sharing personal information.	Regression	Feedback type (immediate/delayed, actionable/ non-actionable)	For non-actionable, immediate and delayed feedbacks differed significantly.
[7]	Predicting interruptibility with GPS and calendar	Three users Not stated	Participants rated their interruptibility level. The scale was not specified in detail.	Descriptive statistics	None	The study computed interruptibility based on GPS and calendar, then participants rated their perception about interruptibility. The computed and rated interruptibility levels did not match.
[143]	Aims to predict available moments for sending notifications	Six users Two months	User feedback was retrieved with a simple question: "Is it a good time?"	Naïve Bayes	Activity, location, time, application usage, ringer mode, and media usage	Accuracy ranges between 73.5% - 90.7% from user to user

Continued on next page.

Study	Objective/ Hypothesis	No of participants Data collection period	Scales used	Method	Significant features	Result
[75]	Understanding people's receptivity to mobile notifications	20 users Two months	Seven multiple-choice and two free-response questions about disruptiveness, acceptance or dismission of notifications were asked.	ANOVA, t-test Kendall's Tau correlation, logistic regression, linear regression	Alert type, priority of the notification	Relationship with the sender of the notification, continuing task type, completion level, complexity are effective on perceived disruption. Alert type is also effective on response time in addition to the parameters above.
[144]	Understanding user's preferences for receiving notifications based on rule extraction	16 subjects 15 days	No scales used	Association rule mining	Notification type and location	They explained the circumstances that users dismissed a notification with a precision of 91%.
[182]	Interruptibility prediction with a longer history of sensors	25 users Four weeks	The participants stated whether they are interruptible or not.	SVM, Random Forest	Accelerometer, light, application category, time, screen on/off, battery temperature, battery level, mobile connectivity	Considering the beginning of the current day for interruptibility prediction reached a 90% accuracy.
[183]	Understanding of interruptions in multi-device environments	16 participants One week	Questionnaire about each device's (smartphone, tablet, PC, smartwatch) proximity and whether users prefer to receive notification on each device.	Friedman test, Wilcoxon test, correlation	Proximity to device, whether the device is currently being used, location	Participants prefer to be notified on smartphones.
[170]	Measuring how notifications interrupt users at specific locations	68 participants Not specified	5-level Likert scale for assessing interruptibility for each location type	Descriptive statistics	Location	Location type of the users affects their interruptibility.

Continued on next page.

Study	Objective/ Hypothesis	No of participants Data collection period	Scales used	Method	Significant features	Result
[172]	Detecting user's breakpoints	30 participants 16 days	NASA-TLX questionnaire	Friedman test, Wilcoxon test	Transition between mobile applications	Frustration scores of the participants whose notifications deferred to breakpoints are lower than the ones whose notifications sent at random times.
[184]	Investigating interruptibility based on activity breakpoints	30 participants 4 days	5-level Likert scales for assessing stress, busyness and tiredness	Descriptive statistics	Activity breakpoints	Notifications sent at activity breakpoints increase the response rate and time
[185]	Investigating whether use of smartphones and walking have an effect on work interruptibility	7 participants 111 days	User rating about interruptibility	ANOVA	Typing activity, PC operation activity	Interruptibility is significantly lower when workers have a typing activity and a PC activity.
[34]	Reducing unintended interruptions	48 participants 4 weeks	Questionnaire about demographic and general app use information	Random forest, k-means	Physical activity, continuing session, dialog generation delay, network type	81.7% accuracy on user interaction, 75.5% of interruptions can be prevented and delayed.
[169]	Predicting user's reachability, engage-ability and receptivity	224 participants 178 days	None	Mann-Whitney Test Kruskal-Wallis Test C4.5 classifier	Acceleration, light, screen coverage, ringer mode, orientation, charging state, time and day of week	Predictive models can be built with >80% precision for most users.

Continued on next page.

Study	Objective/ Hypothesis	No of participants Data collection period	Scales used	Method	Significant features	Result
[73]	Inferring user engagement with the content of notifications	37 users 4 weeks	Scale for assessing the mood of the user	XGBoost	Time passed since the last phone call and SMS, the number of notifications received in the day, physical activity, distance to work, time spent at home, noise, light, age, time, battery charge, app use	The intelligent system, which uses context data achieved 66.6% higher success rate than baseline.
[18]	Investigating interruptibility of office workers using computer interaction and biometric sensors	13 participants 2 weeks	7-level Likert scale for measuring interruptibility	Random forest	Time windows of features, computer monitoring, biometric sensor data	Time windows of more than 5-min give more accurate results on interruptibility. Biometric sensors and computer monitoringgive the most accurate results for inferring interruptibility.
[76]	Predicting interruptibility level of users	33 participants 4 weeks	Scales for personality, mood and interruptibility level	SVM, decision tree, additive regression	Personality, mood location, relation with the interrupter and activity	The model reached 66.1% accuracy for predicting interruptibility level.
[118]	Understanding the receptivity for mobile health interventions	31 participants 3 weeks	Questionnaire for availability, location, social setting, level of focus, level of fatigue, ongoing task	Multilevel logistic regression	Ongoing task, level of focus, social setting, location	Ongoing task, social setting, level of focus are effective on perception of interventions. Ongoing task, social setting, level of focus, location are effective on availability.

A.2 Message Samples for Digital Health Interventions

Study	Target Domain	Medium	Message Content
[186]	Physical activity promotion for diabetic subjects	E-health platform	**Motivational Stage:** "Did you know that a regular and constant physical exercise makes the heart more resistant to possible ischemia?" **Performance Level:** "Very good! You're maintaining your performance!" **Emotional Status:** "Excellent performance! Probably walking helps you to relieve your stress" **Progress along the exercise path:** "Don't be discouraged! Changing habits is a process that evolves along time: the difficult part is most all at the beginning." **Location:** "The Weather Forecast says it will be probably raining this weekend. Why don't you go and visit an exposition near your town?"
[90]	Physical activity promotion for postnatal women	SMS	**Self-efficacy:** "Talk to X about watching the kids while you exercise. You could set a regular time each week, so plans are in place." **Goal setting skills:** Your treat for reaching this weeks' exercise goal is a bubble bath. It's a treat you deserve, so work for it." **Social support:** "Make a deal with Y to watch the kids while you do exercise & then return the favor." **Perceived environmental opportunity for activity:** "Free walking group for mums starts Mon 25th June at 9:30 AM in Apex Park near the lake. Prams welcome. Join the group."
[91]	Prolonged sitting and taking breaks	Computer	**Positive reminder:** "Moving helps with creativity. Take a short walk around the office to help yourself a difficult problem" **Feedback:** "You've taken 5 breaks today! Keep up the good work!"

Study	Target Domain	Medium	Message Content
[30]	Prolonged sitting and active breaks	SMS	**Authority:** "WHO advises to be active on a daily basis. Being inactive for prolonged periods is bad for your health." **Commitment:** "You have already been using the activity monitor. Keep active to reach your daily goals." **Consensus:** "Get off your chair and move! 95% of the participants have already increased their physical activity. Follow their example!" **Scarcity:** "Every day without physical activity is a missed chance to reach a healthier life. Stay active!"
[94]	Activity improvement for COPD	Mobile application	**Encouraging:** "You took more rest, we advise you to take a short walk." **Discouraging:** "You have been very active, take some time to read a magazine" **Neutral:** "You are doing well, keep up the good work"

APPENDIX B

UBDROID MATERIAL

B.1 Web and Client Screenshots

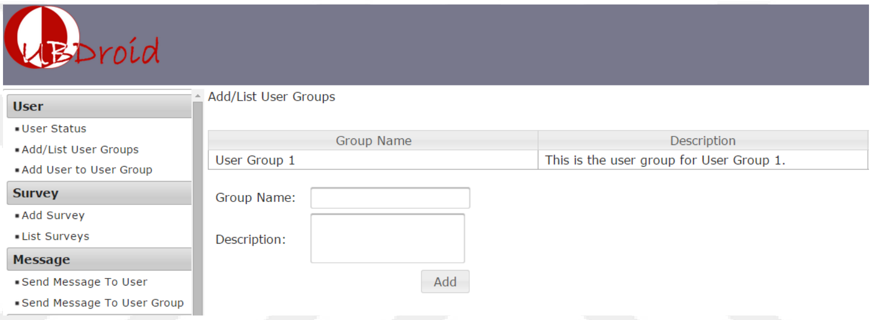


Figure 20: Web interface of UBDroid for creating a new user group

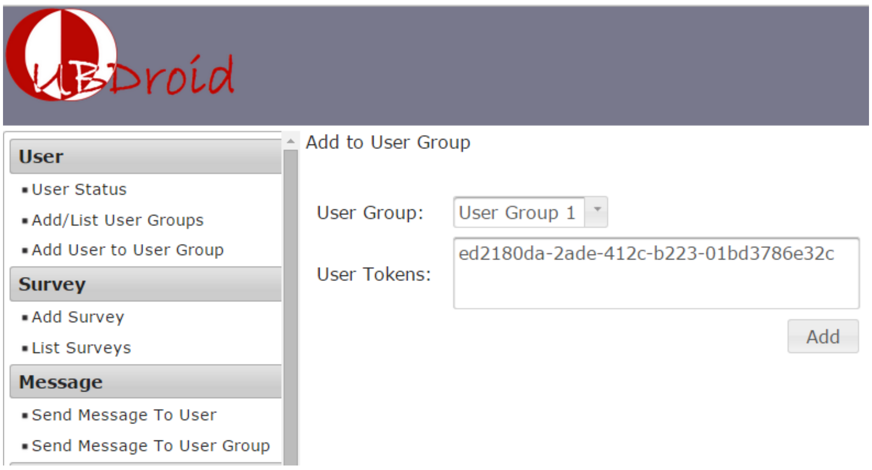


Figure 21: Web interface of UBDroid for adding a user to a user group

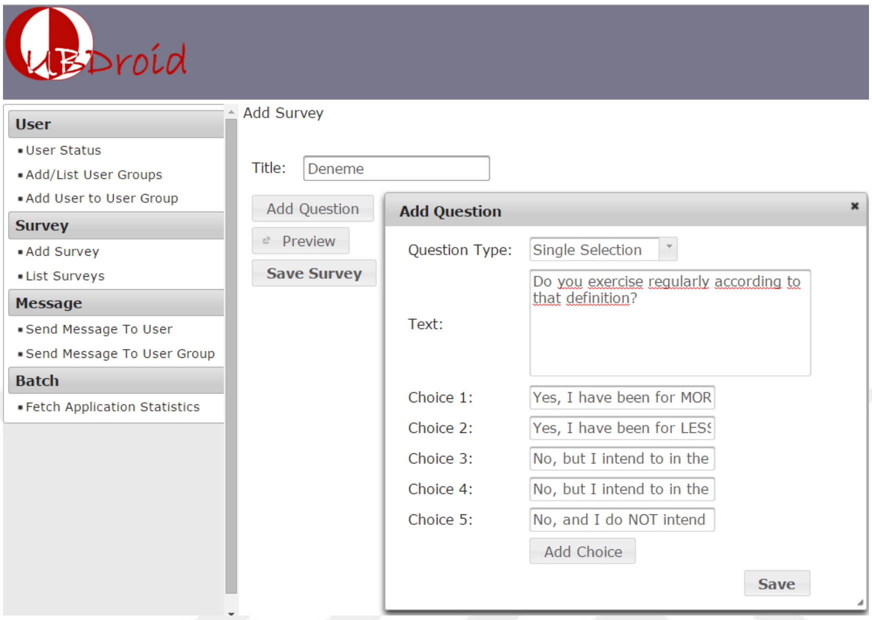


Figure 22: Web interface of UBDroid for creating a new survey

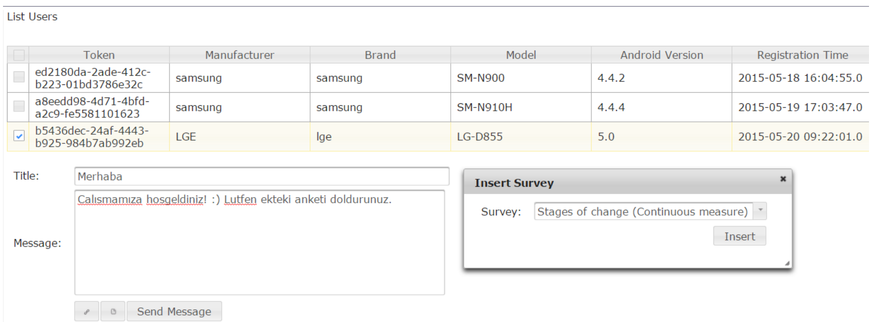


Figure 23: Web interface of UBDroid for sending a message to a user

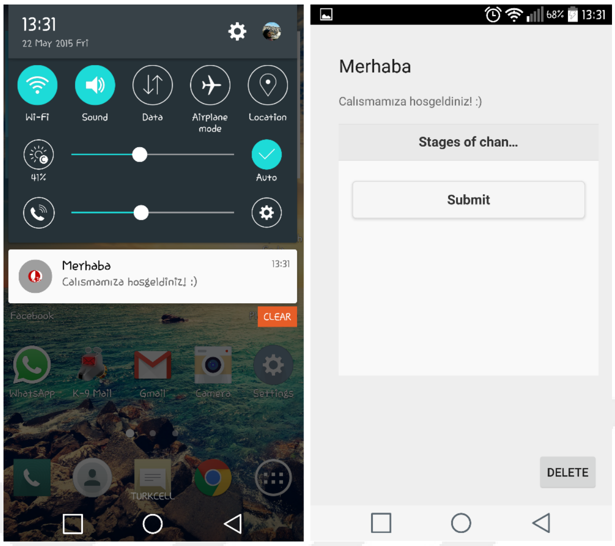


Figure 24: (Left) Message delivery as a notification on mobile phone, (Right): Message appearance after tapping notification

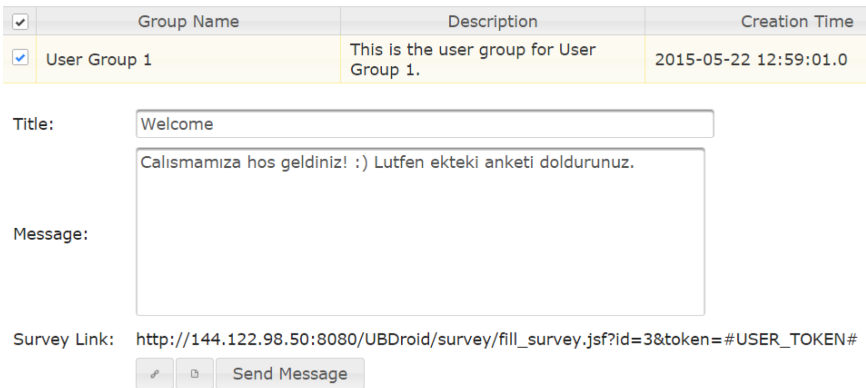


Figure 25: Web interface of UBDroid for sending a message to a user group

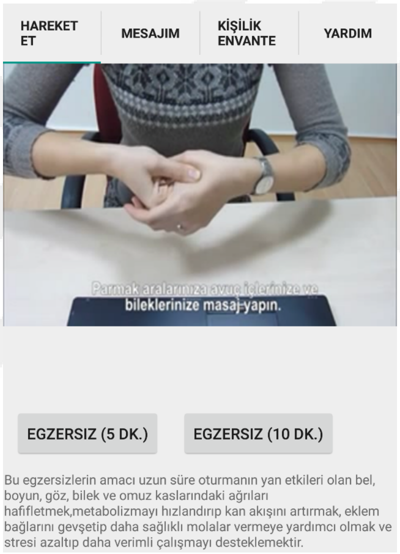
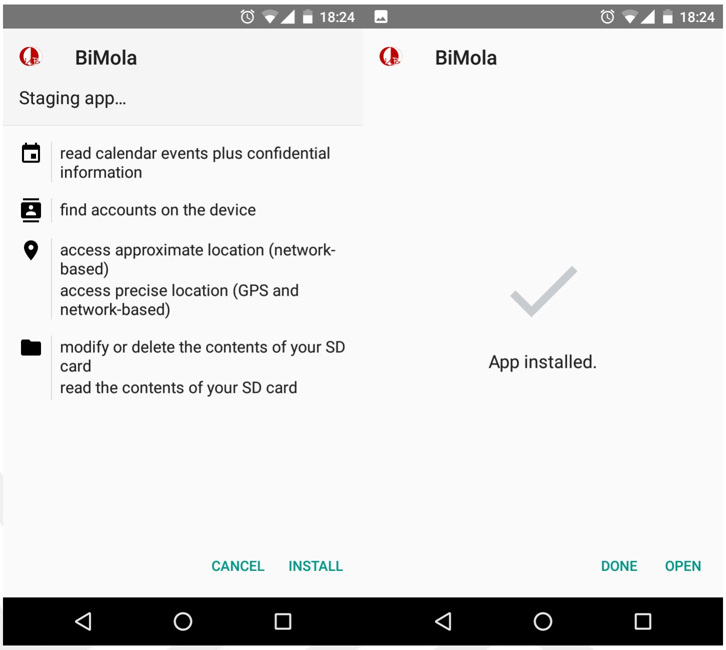
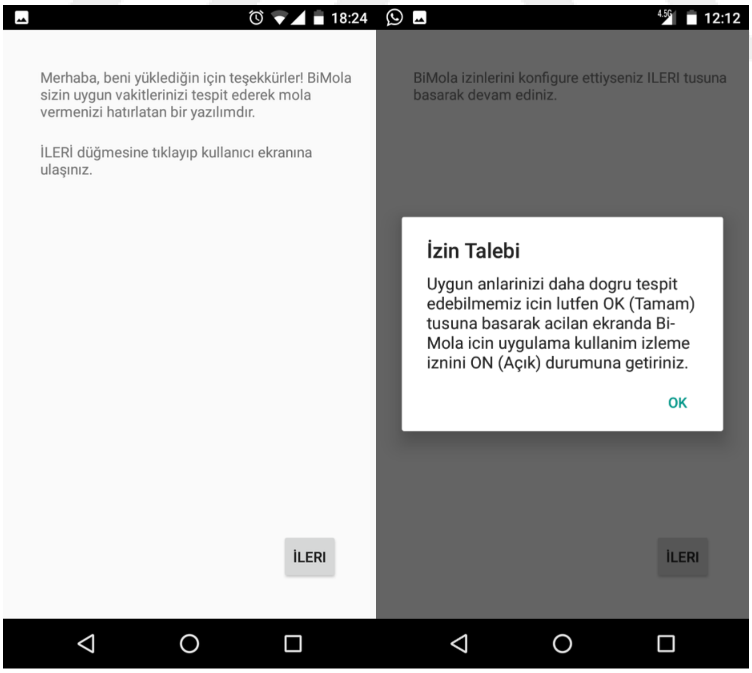


Figure 26: Exercise videos on UBDroid

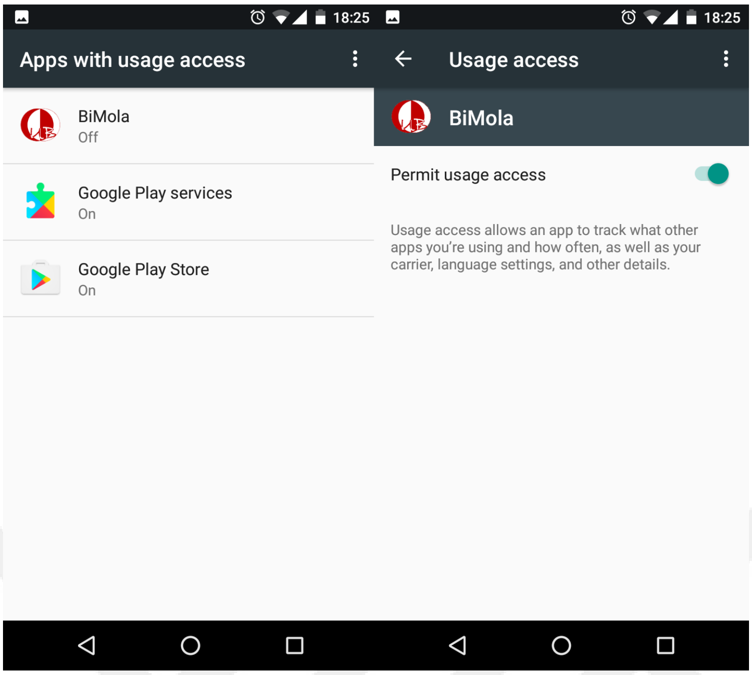
B.2 Successful Installation Steps of UBDroid



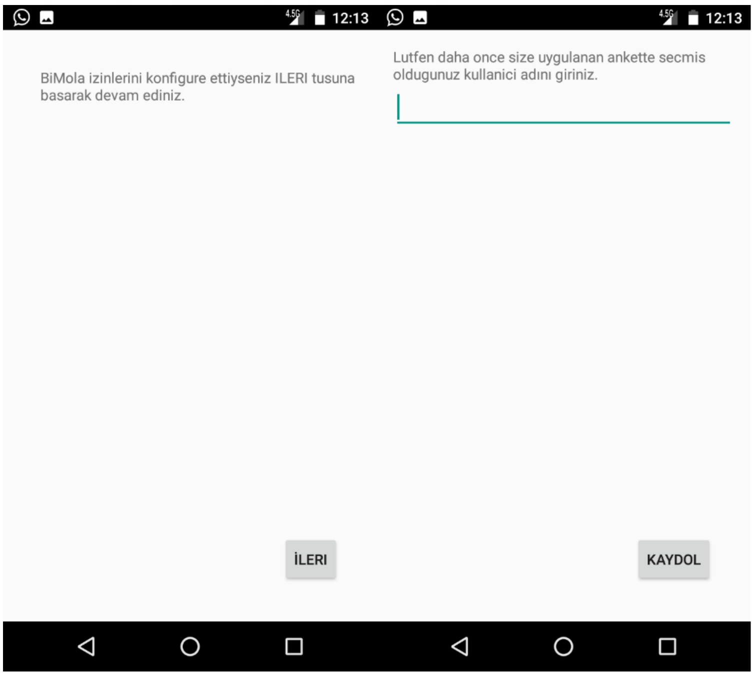
Step 1 (left) and Step 2 (right)



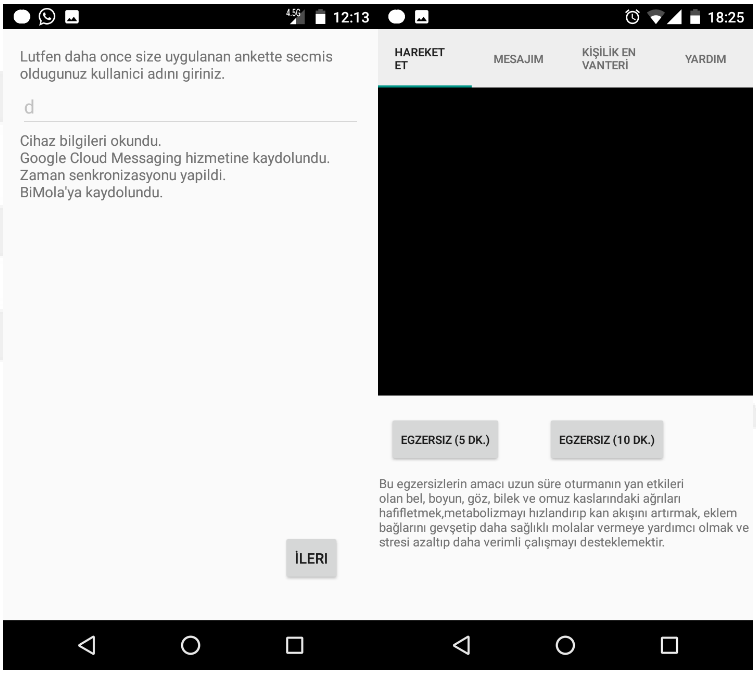
Step 3 (left) and Step 4 (right)



Step 5 (left) and Step 6 (right)

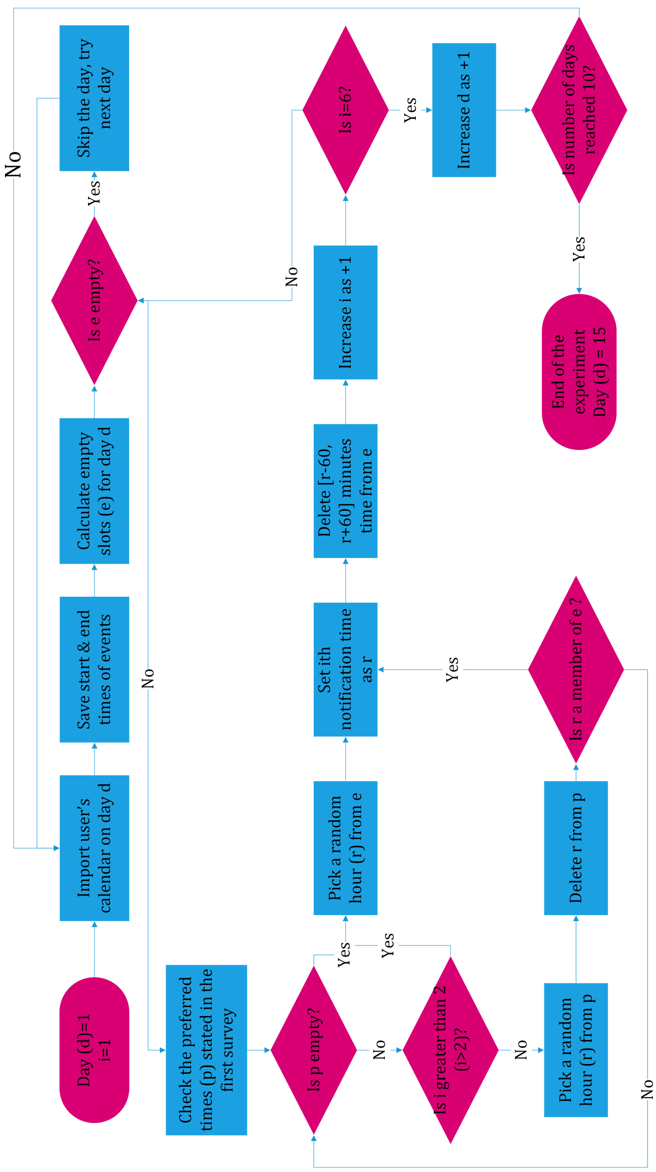


Step 7 (left) and Step 8 (right)



Step 9 (left) and Step 10 (right)

B.3 Steps of the Notification Delivery Algorithm



APPENDIX C

INSTRUMENTS, ETHICAL CLEARANCE AND EXPERIMENT LEAFLET

C.1 Pre-Experiment Questionnaire

1. Yaşınız?
2. Cinsiyetiniz?
   - Kadın
   - Erkek
3. Mesleğiniz?
4. Çalıştığınız kurum?
   - Özel bir kuruluşta çalışıyorum.
   - Bağımsız (freelance) olarak çalışıyorum.
   - Bir kamu kuruluşunda çalışıyorum.
   - Şirket ortağıyım/sahibiyim.
   - Diğer (Belirtiniz) ......
5. Çalıştığınız kurumun alanı?
   - Bilişim
   - İletişim/Haberleşme
   - Savunma sanayi
   - Danışmanlık
   - Eğitim
   - Banka/Finans
   - Sigortacılık
   - Sağlık
   - Enerji
   - Gıda
   - Otomotiv
   - Ağır sanayi (Demir çelik vb.)
   - Diğer (Belirtiniz) ......

6. İşyerinizin büyüklüğü?

   - Mikro ölçekli işletme (Çalışan sayısı 10'dan az, ciro 2 milyon euro'ya kadar)
   - Küçük ölçekli işletme (Çalışan sayısı 50'den az, ciro 10 milyon euro'ya kadar)
   - Orta ölçekli işletme (Çalışan sayısı 250'den az, ciro 50 milyon euro'ya kadar)
   - Büyük ölçekli işletme (Çalışan sayısı 250'den fazla, ciro 10 milyon euro'dan fazla)

7. İşyerinde çalıştığınız pozisyon?

   - Mühendis
   - Uzman
   - Şef
   - Orta Düzey Yönetici
   - Üst Düzey Yönetici
   - Öğretim Üyesi/Görevlisi
   - Araştırma Görevlisi
   - Diğer (Belirtiniz) ......

8. Ne kadar süredir bu pozisyonda çalışıyorsunuz? ..... Yıl .... Ay

9. Günlük bilgisayar kullanımınız ne kadardır?

   - 2 saatten az
   - 2-4 saat
   - 4-6 saat
   - 6-8 saat
   - 8 saatten fazla

10. İş yerinde bulunmanız gereken ofis saatleri nedir? Başlangıç (ss:dd)...... Bitiş (ss:dd):.......

11. Ofis ortamınızda siz dahil kaç kişi çalışıyor?

12. Ofis ortamında çalışırken aşağıdaki kurallardan hangisi/hangileri geçerli?

    - Sesli konuşmak uygun değil.
    - Telefon getirmek yasak.
    - Telefon sessizde olmak durumunda.
    - Telefon kapatılmak durumunda.
    - Herhangi bir kısıtlama yok.

13. Bulunduğunuz ofis ortamında egzersiz, spor yapanlar var mı?

    - Evet
    - Hayır

14. İşiniz cep telefonunu kullanmayı ne kadar gerektiriyor? [1: Hiç gerektirmiyor] [5:Çok gerektiriyor]

15. Çoğunlukla **düzenli** olarak verdiğiniz ara saatlerini ve yerlerini giriniz. Bu mola zamanlarının hangilerinin başında/sonunda/ortasında egzersiz yapmak için uygun olup olmadığınızı belirtiniz. Mola saatlerini girerken yaklaşık başlangıç saati ve bitiş saatini belirtiniz. Özel bir gün belirtecekseniz birden fazla olanları belirtip noktalı virgülle ayırınız (Örneğin: Cuma; Pazartesi). Mekan olarak ise genelde gittiğiniz mekan ile bilgi giriniz. Verdiğiniz ara birden fazla sıralı aktiviteyi içerebilir. Mola tipi olarak uygun olanları birden fazla olarak seçebilirsiniz. Aynı şekilde mola yerlerini de birden fazla seçebilirsiniz.

	Ara 1		Ara 2	
Mola saatleri ve günleri	Başlangıç saati ve dakikası	ss : dd	Başlangıç saati ve dakikas	ss : dd
	Bitiş saati ve dakikası	ss : dd	Bitiş saati ve dakikası	ss : dd
	Günü	☐ Her gün ☐ Her günü yazınız	Günü	☐ Her gün ☐ Her günü yazınız
Mola Tipi	Çay/kahve molaları	☐	Çay/kahve molaları	☐
	Sigara molaları	☐	Sigara molaları	☐
	Yemek ve atıştırma molaları	☐	Yemek ve atıştırma molaları	☐
	Sosyalleşme için ayırdığım molalar (telefonla ya da sosyal çevre ile sohbet)	☐	Sosyalleşme için ayırdığım molalar (telefonla ya da sosyal çevre ile sohbet)	☐
	Dini sebeplerle ilgili ayırdığım molalar	☐	Dini sebeplerle ilgili ayırdığım molalar	☐
	Sağlık sebebiyle ayırdığım molalar (dinlenme, egzersizler, ilaç veya tedavi ile	☐	Sağlık sebebiyle ayırdığım molalar (dinlenme, egzersizler, ilaç veya tedavi	☐
	İhtiyaç molaları (tuvalet, makyaj, vs.)	☐	İhtiyaç molaları (tuvalet, makyaj, vs.)	☐
	Gündemi takip etme molaları (gazete, dergi, vs. okuma faaliyetleri)	☐	Gündemi takip etme molaları (gazete, dergi, vs. okuma faaliyetleri)	☐
	Diğer (1)	☐	Diğer (1)	☐
	Diğer (2)	☐	Diğer (2)	☐
	Diğer (3)	☐	Diğer (3)	☐
	Diğer (4)	☐	Diğer (4)	☐
Mola Yeri	Aynı ofiste	☐	Aynı ofiste	☐
	Bina dışında	☐	Bina dışında	☐
	Ofis dışı, aynı binada farklı katta	☐	Ofis dışı, aynı binada farklı katta	☐
	Ofis dışı aynı katta	☐	Ofis dışı aynı katta	☐
Egzersiz Uygunluğu	Evet	☐	Evet	☐
	Hayır	☐	Hayır	☐

16. 10 dakikalık ufak bir egzersiz molası (yürüyüş, basit boyun-bilek-göz hareketleri vb.) vermeyi planlıyorsanız **yukarıda belirttiğiniz dışında** uygun olduğunuz zamanların yaklaşık başlangıç saatini (ss:dd şeklinde) ve süresini (dakika olarak) belirtiniz. Birden fazla ise noktalı virgül ile birbirinden ayırınız (Örneğin: başlangıç saatleri 10:00;16:00 ve süreleri de 30;10 olarak yazdığınızda, saat 10:00'da 30 dakikalığına, saat 16:00'da 10 dakikalığına ara verebileceğinizi gösterir).
    Başlangıç saati (ss:dd olarak giriniz) .... Süresi (dakika olarak giriniz) ....

17. Molalarda cep telefonunuzun ses modu ve konumu (yanınızda olup olmama durumu) **çoğunlukla** hangi durumda olur? Lütfen size uyan mola tipleri için işaretleyiniz.

	Telefonun modu				Telefonun konumu	
	Titreşim	Sesli	Sessiz	Telefon Kapalı	Yanımda olur	Yanımda olmaz
Çay/kahve molaları						
Sigara molaları						
Yemek ve atıştırma molaları						
Sosyalleşme için ayırdığım molalar(telefonla ya da sosyal çevre ile sohbet)						
Dini sebeplerle ilgili ayırdığım molalar						
Sağlık sebebiyle ayırdığım molaları (egzersizler, ilaç ve tedavi ile ilgili)						
İhtiyaç molaları (tuvalet, makyaj, vs.)						
Gündemi takip etme molaları (gazete, dergi vs. okuma)						
Diğer (1)________						
Diğer (2)_________						
Diğer (3)________						
Diğer (4)						

18. Aşağıdaki durumlarda cep telefonunuzun ses modu ve konumu (yanınızda olup olmama durumu) **çoğunlukla** hangi durumda olur? Lütfen size uyan mola tipleri için işaretleyiniz.

	Telefonun modu				Telefonun konumu	
	Titreşim	Sesli	Titreşim	Sesli	Titreşim	Sesli
Restoran, cafe vb.						
Arkadaşımın yanındayken						
Toplantıda						
İş saatlerinde ofiste						
Evde						
Uyurken						
Sinema, tiyatro vb.						
Ailemle birlikteyken						
Yolda (araba, otobüs vb.)						

19. Bilgisayar başında çalışırken belirli periyotlarda ara veriyor musunuz?
    - Evet, düzenli aralıklarla ara veriyorum.
    - Ara vermeye çalışıyorum ancak unutuyorum/işime dalıyorum.
    - Mümkün olduğunca ara vermeden işime konsantre olmayı tercih ediyorum.
20. Bilgisayar başında el, bilek ya da omuzlarınız için esneme ya da egzersiz hareketleri gerçekleştiriyor musunuz? (1: Hiç gerçekleştirmiyorum, 5: Her zaman gerçekleştiriyorum)
21. Bilgisayar kullanırken el, bilek ya da omuzlarınızda ağrı ya da uyuşukluk hissediyor musunuz? (1: Hiç hissetmiyorum, 5: Çok sık hissediyorum)

22. Daha evvel ergonomi ile ilgili bir doktor tarafından teşhis kondu mu? Örneğin; karpal tünel sendromu, boyun fıtığı vb.
    - Evet
    - Hayır
23. Rahatsızlığınız ile ilgili bir tedavi aldınız mı? Örneğin; fizik tedavi, ilaç tedavisi, cerrahi operasyon vb.
    - Evet
    - Hayır
24. Tedavi süresince ya da sonrasında egzersiz verildi mi?
    - Evet
    - Hayır
25. Tedavinizdeki egzersizleri düzenli gerçekleştirdiniz mi?
    - Programa tamamen sadık kaldım.
    - Çoğunlukla gerçekleştirdim.
    - Bazen gerçekleştirdim.
    - Çok az gerçekleştirdim.
    - Hiç gerçekleştirmedim.
26. Tedavi sürecinizi nasıl sonlandırdınız?
    - Tedavim devam ediyor.
    - Tedavim doktorumun önerisiyle sona erdi.
    - Tedavimi kendim yarıda bıraktım.
27. Lütfen aşağıdaki ifadelere ne derece katıldığınızı 1-5 arasında oylayarak belirtiniz.
    [1: Hiç katılmıyorum, 5: Tamamen katılıyorum]
    (a) Eğer önemli engeller olmazsa, iş saatlerimde kısa dinlenme araları vermeyi düşünürüm.
    (b) Fırsatım olursa, kısa dinlenme araları vermek isterim.
    (c) Kısa dinlenme araları, etkinliğimi artırır.
    (d) Kısa dinlenme araları, işteki performansımı artırır.
    (e) Kısa dinlenme araları, iş kalitemi artırabilir.
    (f) Ara vermeyi hatırlatıcı uygulamaları kullanmak, ofis ortamındaki sağlık ihtiyaçlarımı karşılayacaktır.
    (g) Ara vermeyi hatırlatıcı uygulamaları kullanmak, ofis ortamındaki sağlığımı yönetme etkililiğimi artıracaktır.
    (h) Genel olarak, ara vermeyi hatırlatıcı uygulamaları kullanmak, ofis sağlığımı yönetmede faydalı olacaktır.
    (i) Ara vermeyi hatırlatıcı uygulamaları kullanmak, ofis sağlığımı yönetmek için güzel bir fikir.
    (j) Ara vermeyi hatırlatıcı uygulamaları kullanma fikri hoşuma gidiyor.

(k) Ara vermeyi hatırlatıcı uygulamaları kullanmak, ofis sağlığımı yönetmek için akıllıca bir fikir.

(l) Ara vermeyi hatırlatıcı uygulamaları kullanmak, ofis sağlığımı yönetmek için değerli olacaktır.

28. Orta düzeyde fiziksel aktiviteler nefes alımında ve kalp atımında biraz artış gözlenen aktivitelerdir. Ritimli yürüyüş, dans, bahçe işleri, düşük şiddette yüzme veya arazide bisiklet sürme gibi etkinlikler orta düzeyde aktivite olarak değerlendirilir.
Orta düzeyde fiziksel aktivitenin düzenli sayılabilmesi için, aktivitenin haftada 5 veya daha fazla günde 30 dakika veya daha fazla olması gerekir. Örneğin, 30 dakika süreyle yürüyüş yapabilir veya 10 dakikalık 3 farklı aktivite ile 30 dakikayı doldurabilirsiniz.
Lütfen her soru için Evet veya Hayır seçeneğini işaretleyiniz.

	Evet	**Hayır**
Şu anda orta düzeyde fiziksel aktiviteye katılmaktayım.		
Gelecek 6 ayda orta düzeyde fiziksel aktiviteye katılımımı artırmak niyetindeyim.		
Şu anda düzenli olarak orta düzeyde fiziksel aktivite yapmaktayım.		
Son 6 aydır düzenli olarak orta düzeyde fiziksel aktiviteye katılmaktayım.		
Geçmişte, en az 3 aylık dönemde düzenli olarak orta düzeyde fiziksel aktivitelere katılırdım.		

29. Son 30 gündür herhangi egzersizle ilgili bir takip edici cihaz ya da uygulama kullanıyor musunuz? (Örneğin; adım sayarlar, Runtastic, Google Fit vb.)

- Evet
- Hayır

30. Egzersiz yapmayı hatırlatan uygulamaların gönderdiği mesajların zamanlaması hakkında ne düşünüyorsunuz? [1:Çok Kötü][5:Çok iyi]

31. Lütfen aşağıdaki ifadelere ne derece katıldığınızı 1-5 arasında oylayarak belirtiniz.
[1: Hiç katılmıyorum, 5: Tamamen katılıyorum]

(a) Yöneticilerim dinlenme molası verip vermeme kararımı etkiler.

(b) İş arkadaşlarım dinlenme molası vermem için bana destek olur.

(c) Beni etkileyen insanlar/çevrem ara vermeyi hatırlatıcı uygulamaları kullanmamı söyler.

(d) Benim için önemli olan kişiler ara vermeyi hatırlatıcı uygulamaları kullanmam konusunda beni teşvik eder.

(e) Kurum yönetimi ara vermeyi hatırlatıcı uygulamaları kullanımımı destekler.

32. Günlük işlerinizi, toplantılarınızı takip etmek için **klasik takvim (kağıt üzerinde tutulan ajanda tipi takvimler)** kullanıyor musunuz?

- Evet
- Hayır

33. **Klasik takviminizi** güncelleme durumunuz nedir?

- Her toplantımı, işimi vb. takvime düzenli olarak girerim.

- Yalnızca önemli ya da unutmamam gereken olayları takvime girerim.
- Takvimi arada güncellerim.
- Hiç güncellemem.

34. **Klasik takviminize** gün içindeki iş ve toplantılarınızı girmeye ne derece önem veriyorsunuz? (1: Hiç önem vermem, 5: Çok önem veririm)
35. Aşağıdaki etkinliklerden hangilerini **klasik takviminize** giriyorsunuz? (Birden fazla işaretleyebilirsiniz)

    - İş ile ilgili etkinlikler
    - Özel hayat etkinlikleri (aile, arkadaşlar ile yapılan etkinlikler)
    - Aile bireylerine ait etkinlikler (Örn; eşinizin toplantıları, çocuğunuzun kurs saatleri vb.)
    - Tatil programları
    - Doğum günleri/yıl dönümleri
    - Diğer (Belirtiniz) ....

36. Günlük işlerinizi, toplantılarınızı takip etmek için **dijital takvim (Outlook, Google Calendar vb.)** kullanıyor musunuz?

    - Evet
    - Hayır

37. **Dijital takviminizi** güncelleme durumunuz nedir?

    - Her toplantımı, işimi vb. takvime düzenli olarak girerim.
    - Yalnızca önemli ya da unutmamam gereken olayları takvime girerim.
    - Takvimi arada güncellerim.
    - Hiç güncellemem.

38. **Dijital takviminize** gün içindeki iş ve toplantılarınızı girmeye ne derece önem veriyorsunuz? (1: Hiç önem vermem, 5: Çok önem veririm)
39. Aşağıdaki etkinliklerden hangilerini **dijital takviminize** giriyorsunuz? (Birden fazla işaretleyebilirsiniz)

    - İş ile ilgili etkinlikler
    - Özel hayat etkinlikleri (aile, arkadaşlar ile yapılan etkinlikler)
    - Aile bireylerine ait etkinlikler (Örn; eşinizin toplantıları, çocuğunuzun kurs saatleri vb.)
    - Tatil programları
    - Doğum günleri/yıl dönümleri
    - Diğer (Belirtiniz) ....

40. Dijital takvim **kullanmaMA** sebebiniz nedir? (Uygun olan seçenekleri işaretleyebilirsiniz)

    - Yapacağım işleri ve toplantıları hatırlayabiliyorum.
    - Kullanmaya ve bilgi girmeye üşeniyorum.
    - Klasik takvim kullanmayı tercih ediyorum.
    - Güvenlik/mahremiyet nedeniyle kullanmıyorum.

- Kullanmakta zorlanıyorum.
- Pratik bulmuyorum.
- Kullanmayı sevmiyorum.
- Otomatik hatırlatlamalarından rahatsız oluyorum.
- Diğer (Belirtiniz) ....

C.2 Basic Personality Traits Inventory (BPTI)

Aşağıda size uyan ya da uymayan pek çok kişilik özelliği bulunmaktadır. Bu özelliklerden her birinin sizin için ne kadar uygun olduğunu 1 ile 5 arasında belirtiniz (1:Hiç uygun değil, 5: Tamamen uygun). Örneğin;

Kendimi ……….. biri olarak görüyorum.

1. Aceleci
2. Yapmacık
3. Duyarlı
4. Konuşkan
5. Kendine güvenen
6. Soğuk
7. Utangaç
8. Paylaşımcı
9. Geniş/rahat
10. Cesur
11. Agresif
12. Çalışkan
13. İçten pazarlıklı
14. Girişken
15. İyi niyetli
16. İçten
17. Kendinden emin
18. Huysuz
19. Yardımsever
20. Kabiliyetli
21. Üşengeç
22. Sorumsuz
23. Sevecen
24. Pasif
25. Disiplinli
26. Açgözlü
27. Sinirli
28. Cana yakın
29. Kızgın
30. Sabit fikirli
31. Görgüsüz
32. Durgun
33. Kaygılı
34. Terbiyesiz
35. Sabırsız
36. Yaratıcı
37. Kaprisli
38. İçine kapanık
39. Çekingen
40. Alıngan
41. Hoşgörülü
42. Düzenli
43. Titiz
44. Tedbirli
45. Azimli

C.3 BPTI Scores of the Participants

User No	EXT	CONS	AGR	NEU	OPN	NV
U01	3.25	3.75	3.63	2.88	2.00	2.67
U02	2.75	3.63	4.13	2.75	2.38	1.33
U03	2.13	3.63	4.50	2.50	2.13	1.17
U04	3.13	3.88	3.88	3.63	2.38	2.50
U05	3.63	3.50	4.00	2.38	2.88	2.50
U06	3.50	4.13	4.38	2.88	2.88	1.50
U07	3.13	2.50	4.25	3.38	2.50	1.17
U08	3.50	1.38	4.00	2.13	2.75	1.00
U09	3.25	3.38	4.38	4.25	2.63	1.50
U10	2.63	2.50	3.88	2.88	2.13	1.67
U11	2.13	3.00	3.88	3.88	2.00	2.17
U12	4.00	3.75	4.50	3.63	3.00	1.33
U13	2.63	3.88	4.75	3.25	2.00	2.50
U14	2.38	4.38	4.38	3.50	3.00	1.33
U15	3.50	4.00	5.00	3.00	2.50	1.00
U16	3.00	3.88	4.38	3.25	2.88	1.83
U17	4.63	4.00	4.75	2.75	3.38	1.00
U18	4.25	3.63	4.00	2.75	3.00	1.50
U19	3.00	2.75	4.00	4.13	2.63	1.50

EXT: Extroversion, CONS: Conscientiousness, AGR: Agreeableness, NEU: Neuroticism, OPN: Openness, NV: Negative Valence.

C.4 Mobile Phone Problem Use Scale (MPPUS)

Altta verilen ifadelerin size ne kadar uygun olduğunu aşağıdaki ölçeği kullanarak belirtiniz:

1: Hiç doğru değil, 10: Tamamen doğru

1) Cep telefonumla hiçbir zaman yeteri kadar zaman geçiremiyorum.
2) Kendimi kötü hissettiğimde daha iyi hissetmek için cep telefonumu kullandığım oldu.
3) Başka şeyler yapıyor olmam gerekirken kendimi telefonla uğraşırken buluyorum ve bu sorun yaratıyor.
4) Tüm arkadaşlarım mobil telefon sahibi.
5) Cep telefonumla ne kadar zaman harcadığımı başkalarından gizlemeye çalıştığım oldu.
6) Telefona harcadığım vakit yüzünden uykum kaçıyor.
7) Mobil telefon faturamın bütçemi aştığı olmuştur.
8) Erişim dışı olduğumda, kaçırdığım bir çağrı olacak düşüncesiyle endişeleniyorum.
9) Bazen, mobil telefonumla görüşürken başka işler yapıyorsam, konuşma beni başka yerlere alır götürür ve yaptığım diğer işe dikkat etmem.
10) Cep telefonuyla harcadığım zaman geçtiğimiz 12 ayda arttı.
11) Kendimi dışlanmış (izole olmuş) hissettiğimde, başka kişilerle konuşmak için mobil telefonumu kullandığım olmuştur.
12) Cep telefonumla daha az zaman harcamayı denedim ama başaramıyorum.
13) Mobil telefonumu kapatmak bana çok zor geliyor.
14) Bir müddet telefonumu açmazsam ya da mesajlarımı kontrol etmezsem kendimi huzursuz hissederim.
15) Sık sık cep telefonuyla ilgili rüya görürüm.
16) Ailem ve arkadaşlarım cep telefonu kullanımımdan şikâyet eder.
17) Eğer cep telefonum yoksa arkadaşlarım benimle temasa geçmekte zorlanır.
18) Cep telefonumla harcadığım zamanın direkt bir sonucu olarak üretkenliğim azaldı.
19) Cep telefonu kullanımımla ilişkili olarak ağrı ve sızılarım var.
20) Kendimi, planladığımdan daha uzun süre cep telefonuyla meşgul olurken bulurum.
21) Daha acil diğer işlerle uğraşmak yerine, cep telefonu kullanmayı tercih ettiğim zamanlar var.
22) Uğraşmamam gereken zamanda telefonla uğraştığımdan genellikle randevularıma gecikirim.
23) Toplantıda, yemekte ya da sinemada telefonumu kapatmak zorunda kalırsam sinirli olurum.
24) Cep telefonumla çok fazla zaman harcadığımı söyleyenler oldu.
25) Bir toplantı veya ders sırasında ya da tiyatroda telefonum çaldığı için birkaç kez sıkıntı yaşadım.
26) Arkadaşlarım, cep telefonumun kapalı olmasından hoşlanmazlar.
27) Telefonum olmadan kendimi eksik hissederim.

C.5 ESM Questions

- İşinle ne kadar meşgulsün?
  1: Hiç meşgul değilim, 5: Fazlasıyla meşgulüm
- Yaptığın iş seni ne kadar zorluyor?
  1: Hiç zorlamıyor, 5: Çok fazla zorluyor
- En fazla kaç dakika ara verebilirsin?/Kaç dakikalık mola içindesin?
  - Ara veremem

- 5 dakikadan az
- 5 ile 15 dakika arası
- 15 ile 30 dakika arası
- 30 dakikadan fazla

- Bu vakit hangi egzersizi yapmak için uygun?(Birden fazla seçebilirsiniz).
  - Egzersiz yapmak için uygun zaman değil.
  - Mobil uygulamadaki birinci videodaki hareketleri yapmak için uygun.
  - Mobil uygulamadaki ikinci videodaki hareketleri yapmak için uygun
  - Kendi egzersiz planimi yapmak için uygun.
- Şu anda ne yapıyorsun?
  - Toplantıdayım
  - Bilgisayar (çalışma) ile çalışıyorum
  - Çay/kahve molasındayım
  - Yemek/atıştırma molasındayım
  - Yoldayım
  - Sosyalleşme için ayırdığım moladayım (telefonla ya da sosyal çevre ile sohbet)
  - İhtiyaç molasındayım (tuvalet, makyaj, vs.).
  - Sigara molasındayım
  - Gündemi takip ediyorum (gazete,dergi vs. okuma).
  - Sağlık sebebiyle ayırdığım moladayım (egzersizler, ilaç ve tedavi ile ilgili)
  - Dini sebeple ayırdığım moladayım.
  - Diğer
- Şu anda yaptığınız şeyin başında mı, ortasında mı ya da sonunda mısınız?
  - Başında
  - Ortasında
  - Sonunda
  - Şu an bir şey yapmıyorum.

Questions sent at the end of each experiment day:

- Bugün içinde verilen tavsiyeleri yerine getirdiniz mi?
  - Hiçbirini yerine getirmedim
  - Birazını yerine getirdim
  - Hepsini yerine getirdim
- Bu mesajlar gününüze olumlu bir etki bıraktı mı?
  [1: Olumlu hiçbir etki bırakmadı, 5: Fazlasıyla olumlu etki bıraktı]
- Aldığınız mesajların içeriğini sevdiniz mi?
  [1: Hiç sevmedim 5:Çok sevdim]
- Uygulamayı kullanırken bugünkü motivasyonunuz ne kadardı?
  [1: Hiç motive değildim, 5: Çok fazla motiveydim]

C.6 Post-Questionnaire

Aşağıdaki ifadelere ne derece katıldığınızı 1-5 arasında derecelendiriniz.

1: Kesinlikle katılmıyorum
2: Katılmıyorum
3: Emin değilim
4: Katılıyorum
5: Kesinlikle katılıyorum

1) Bence sunulan uygulama kullanışlıydı.
2) Uygulamayı kullanmak, gönderilen hatırlatmalar bana fayda sağladı.
3) Hatırlatmalar, egzersiz yapmayı daha ilginç hale getirdi.
4) Hatırlatmalar, benzer bilgi sunan geleneksel kaynaklara göre daha iyiydi.
5) Uygulamayı kullanmak netti.
6) Uygulama, kolay kullanılabilirdi.
7) İçerik görüntülemek kolaydı.
8) Uygulamayı kullanmayı ve hatırlatmaları sevdim.
9) Uygulamayı kullanmak ve hatırlatmalar güzeldi.
10) Uygulamayı kullanmak ve hatırlatmalar eğlenceliydi.
11) İleride benzer bir uygulamayı kullanacağım.
12) Uygun olursa, ileride benzer bir sistemi kullanabilirim.
13) Uygulamayı, yakın gelecekte kullanma niyetindeyim.

Uygulamada beğendiğiniz özellikleri belirtiniz.

Uygulamada beğenmediğiniz özellikleri belirtiniz.

Deney süresince size sunulan egzersizleri ofisinizde yapabilme durumunuzu belirtiniz.

- ✓ Hareketleri ofis ortamında yapabildim.
- ✓ Hareketleri ofis ortamında yapamadım. Çünkü ___________________________
- ✓ Hareketleri işyerimde ofis dışında başka bir yerde yapabildim.

Deney süresince size sunulan egzersizlerin sınıflandırması (fark edilmeden yapılabilen, ofiste fark edilerek yapılabilen vb.) sizce uygun muydu?

- ✓ Evet
- ✓ Hayır. Sebebi? __________________________

C.7 Ethical Clearance

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Sayı: 28620816 /103

07 MART 2016

Gönderilen: Doç.Dr. Tuğba Taşkaya TEMİZEL

Enformatik Enstitüsü

Gönderen: Prof. Dr. Canan SÜMER

İnsan Araştırmaları Komisyonu Başkanı

İlgi: Etik Onayı

Sayın Doç.Dr. Tuğba Taşkaya TEMİZEL'in danışmanlığını yaptığı doktora öğrencisi Şeyma ÇAVDAR'ın "İkna Stratejilerindeki Kişisel Farklılıkları belirlemek için mobil kullanıcı verisi işlemesi" başlıklı araştırması İnsan Araştırmaları Komisyonu tarafından uygun görülerek gerekli onay **2016-FEN-006** protokol numarası ile **01.02.2016-31.12.2017** tarihleri arasında geçerli olmak üzere verilmiştir.

Bilgilerinize saygılarımla sunarım.

Prof. Dr. Canan SÜMER

Uygulamalı Etik Araştırma Merkezi

İnsan Araştırmaları Komisyonu Başkanı

Prof. Dr. Meliha ALTUNIŞIK

İnsan Araştırmaları Komisyonu

Üyesi

Prof. Dr. Ayhan SOL

İnsan Araştırmaları Komisyonu

Üyesi

Prof. Dr. Mehmet UTKU

İnsan Araştırmaları Komisyonu

Üyesi

Yrd.Doç.Dr. Pınar KAYGAN

İnsan Araştırmaları Komisyonu

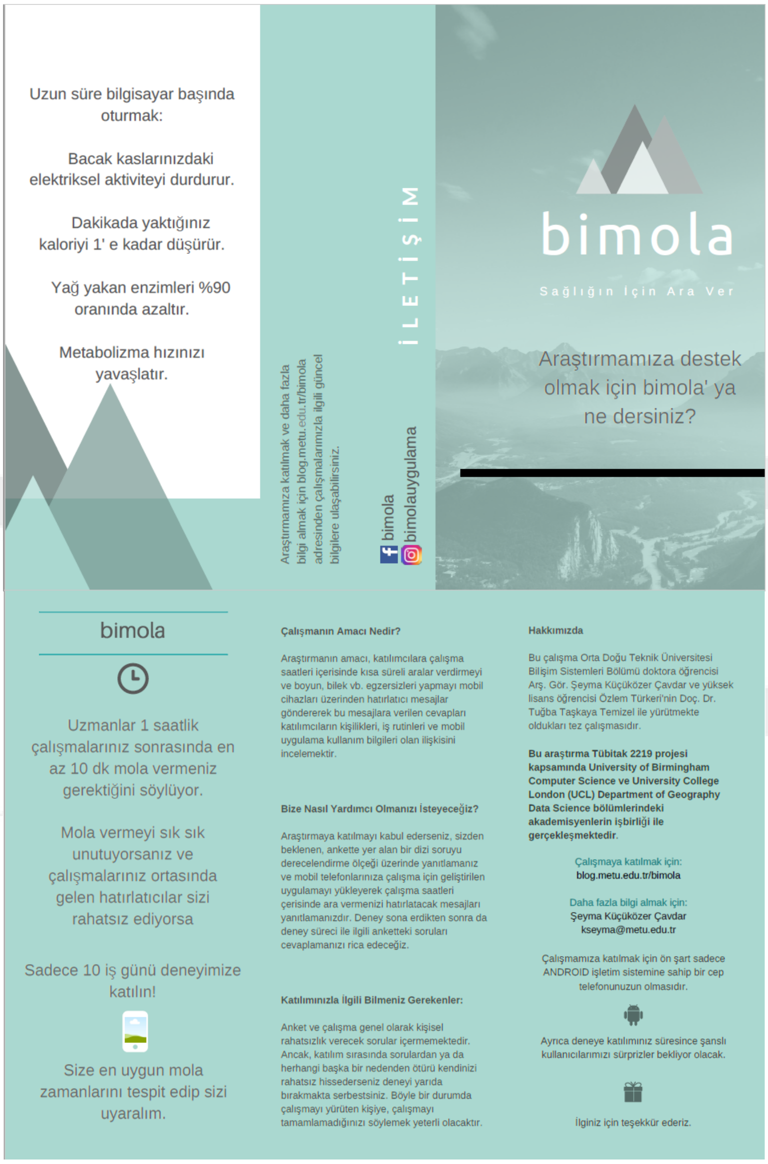
Üyesi

C.8 Motivational Messages Sent with Reminders

- Uzun süre oturmak modern dünyanın sigara alışkanlığı gibidir. Kendine iyi bak ve bir ara ver!
- En iyi dinlenme anı dinlenmeye vakit bulamadığın zamandır.
- Uyarı: Aşırı yüklenme tespit edildi!!! Lütfen bir mola veriniz.
- Bir mola ver ve içindeki sesi dinle.
- Bazen ihtiyacın olan tek şey yaptıklarına bir ara vermektir.
- Rahatlat kendini, sakinleştir zihnini ve yenile ruhunu!
- Kısa bir mola bütün işini değiştirebilir.
- Sakin ol ve bir mola ver!
- Sakinleşmiş bir zihin zorluklar için en büyük silahtır.Bu yüzden bir molanın tam zamanı!
- Gelen kutundaki e-postalar hiç bitmeyecek. O yüzden kafana takma ve kısa bir mola ver.
- Negatif düşünceleri bırak ve bir mola ver!
- Eğer rahatlamayı öğrenirsen ve cevap için beklersen zihnin bir çok soruna cevap verebilir.Rahatlamak için bir mola ver.
- Gerçek performansını göstermek için rahatlamaya ve sakinleşmeye ihtiyacın var bu yüzden bir mola ver.
- Eğer Facebook için zaman bulabiliyorsan, bedenini rahatlatmak için de bulabilirsin.
- Hareket etmek yaratıcılığı geliştirir. Kısa bir yürüyüş zorlu problemlerin üstesinden gelmene yardımcı olacak.
- Derin bir nefes al ve bir mola ver.
- Biraz sakinleşmeyi dene, işler daha çabuk yoluna girecek.
- Bazen büyük resmi görebilmek için bir adım geri atmak gerekir.
- Şu an kendine verebileceğin en güzel hediye küçük bir mola!
- Bazen hiç bir iş yapmadan sakinleşmek işlere doğru perspektifle bakmanı sağlayabilir.
- Meşgul olmak biraz fazla abartılıyor. Sakinleş ve rahatla!
- Meşgul olmak çok da matah bir durum değil.Güzel olan kendine zaman ayırmak ve biraz nefes almak.
- Bu kadar strese girmene gerek yok. Biraz rahatla bir mola ver.
- Tam olarak en meşgul olduğun zamanlar bir molaya en ihtiyaç duyduğun zamanlardır.
- Yenile kendini, bedenini, ruhunu. Sonra işine geri dönersin.
- Çalışmana biraz ara ver. Mola da programının bir parçası!
- Egzersiz yapmak zihnini rahatlatır. Böylece çalışmalarına daha rahat odaklanabilirsin!
- Sıkı çalış, zamanın değerini bil ve bir mola ver!
- Asla vazgeçme ama bir mola ver!
- Hareket Zamanı!
- Rahatla, tazelen ve canlan!

- Eğer zihnini rahatlatabilirsen, zihnin senin için çalışmaya başlar.
- Enerjiniz tükenmek üzere, lütfen mola veriniz.
- Çalışma kaliten ne kadar önemliyse dinlenme kaliten de o kadar önemli!
- Bir günde 1440 dakika var. Kendine zaman ayır ve 10 dakika mola ver.
- Bugün verdiğin molalar sağlığın için çok iyiydi. Bu performansını devam ettir!
- Eğer her saat için kendine mola zamanı yaratabiliyorsan, çalışma hayatının stresiyle daha kolay başa çıkabilirsin.
- Endişelenme, mutlu ol ve bir mola ver!
- Kendinle ilgilenmen için güzel bir zaman.
- Sen önemlisin, kendine zaman ayır ve bir mola ver.
- Rahatlamaya ve yenilenmeye harcadığın tüm zaman kaliteli bir yaşam olarak sana geri dönecek.
- Ortalık karıştıysa çok sorun değil. Bir mola ver ve tekrar devam et.
- Yorgun zihnine yardım et. Bir mola ver ve yeniden başla!
- Eğer sağlığın için kendine zaman yaratmazsan, mola için asla doğru zamanın olmayacak.
- Keyifli bir ana sadece bir mola kadar uzaktasın.
- Az da olsa hareket etmek hiç etmemekten iyidir.
- Sandalyenden kalk ve harekete geç!
- Sadece biraz nefes al.
- Rahatla, şimdi senin zamanın.
- Küçük molalar senin için büyük bir fark yaratabilir
- Bir mola ver. Her şey yoluna girecek...
- Her şeyi kontrol altında tutmak bazen ruhunu bunaltır, kısa bi mola üzerindeki stresi azaltmana yardımcı olur.
- Sadece nefes al. Bu anı bir daha asla yaşayamayacaksın!
- Kendine iyi bir mola için izin ver. Bedeninin buna ihtiyacı var.
- Sakin ol, aradığın herşey bir mola yakınında.
- Bazen çalışmalarının ilerlemesi için gereken tek şey bir moladır.
- Belki programındaki tek eksik verimli bir moladır.
- Sakinleşmek için bir mola ver, çünkü verimli bir mola bedenini ve ruhunu uyandırmana yardımcı olacak.
- Arada bir çalışmalarına bir mola ver ve ruhunu dinle!

C.9 Experiment Leaflet



APPENDIX D

TABLES FOR CHAPTER 6

D.1 DIC Estimates for the GLMMs Used to Predict the Attentional States

Model No	Covariates	Data set 1 $(N = 372)$	Data set 2 $(N = 313)$	Data set 3 $(N = 257)$	Data set 4 $(N = 297)$	Data set 5 $(N = 224)$
1	NOS_{60} + AU_5	503.27±.78	421.26±1.51	337.21±.89	391.30±.79	303.24±1.27
2	NOS_{45} + NOA_{10}	504.54±2.00	423.71±1.16	339.11±1.94	393.32±2.55	300.45±3.19
3	NOS_{45} + AU_5	502.55±1.05	421.78±1.62	335.60±2.49	391.24±3.74	299.49±1.19
4	NOS_{60}	515.37±2.85	431.56±1.45	342.94±3.00	405.28±.87	317.89±1.23
5	NOS_{45}	514.28±2.14	429.65±1.78	340.72±1.45	395.47±2.13	305.88±1.20
6	NOA_{10}	516.82±2.14	430.28±1.98	341.47±.45	398.56±1.76	320.23±.98
7	AU_5	512.56±.87	430.89±1.21	340.21±1.35	396.45±1.45	304.33±.84
8	NOS_{60} + NOA_{10}	507.51±2.41	425.68±1.67	340.57±2.45	394.58±2.14	301.89±1.61
9	NOA_{10} + AU_5	508.57±1.48	423.88±1.78	341.47±1.82	395.69±2.09	302.63±2.39
10	NOS_{60} + AU_5 + NOA_{10}	503.47±.88	424.74±3.10	340.89±1.06	392.71±2.00	300.78±1.83
11	NOS_{45} + AU_5 + NOA_{10}	505.64±1.60	423.97±1.99	341.02±2.06	393.95±1.71	301.11±1.54

Three runs with different window sizes were performed in each sub-sampled data set and the mean and standard deviation of these runs are summarized.

D.2 DIC Estimates for the GLMMs Used to Predict the Engagement Levels

Model No	Covariates	Data set 1 ($N = 372$)	Data set 2 ($N = 313$)	Data set 3 ($N = 257$)	Data set 4 ($N = 297$)	Data set 5 ($N = 224$)
1	NOS_{45} + AU_5	1235.46±3.41	1040.74±1.80	862.34±1.78	949.52±3.71	754.79±1.61
2	NOS_{45} + NOA_{10}	1238.83±1.43	1037.72±3.88	858.34±3.59	953.77±2.05	758.83±1.17
3	NOS_{60} + AU_5	1239.64±1.04	1042.00±2.43	864.96±1.80	951.23±2.83	753.02±4.86
4	NOS_{60}	1248.28±1.21	1050.52±3.89	869.68±2.81	963.56±2.12	773.19±1.73
5	NOS_{45}	1247.19±1.49	1048.61±2.70	867.46±2.31	953.75±3.20	761.18±3.76
6	NOA_{10}	1249.73±1.59	1049.24±1.29	868.21±2.20	956.84±2.18	775.53±1.33
7	AU_5	1245.47±1.25	1049.85±2.20	866.95±1.15	954.73±2.43	759.63±2.39
8	NOS_{60} + NOA_{10}	1240.42±1.87	1044.64±3.26	867.31±2.64	952.86±2.61	757.19±1.69
9	NOA_{10} + AU_5	1241.48±1.39	1042.84±2.06	868.21±4.01	953.97±1.79	757.93±1.43
10	NOS_{60} + AU_5 + NOA_{10}	1236.38±3.89	1043.70±1.49	867.63±2.86	950.99±3.00	756.08±1.67
11	NOS_{45} + AU_5 + NOA_{10}	1238.55±3.27	1042.93±2.67	867.76±1.87	952.23±2.50	756.41±1.77

Three runs with different window sizes were performed in each sub-sampled data set and the mean and standard deviation of these runs are summarized.

D.3 DIC Estimates for the GLMMs Used to Predict the Challenge Levels

Model No	Covariates	Data set 1 ($N = 372$)	Data set 2 ($N = 313$)	Data set 3 ($N = 257$)	Data set 4 ($N = 297$)	Data set 5 ($N = 224$)
1	NOS_{60} + NOA_{10}	1199.97±3.93	1013.63±2.92	821.35±4.55	914.07±2.90	711.49±3.87
2	NOS_{45} + NOA_{10}	1192.73±6.67	1011.47±4.51	824.88±1.90	913.99±5.04	707.31±9.00
3	NOS_{45} + AU_5	1199.25±1.39	1009.10±3.52	820.65±5.54	913.86±6.63	710.52±5.16
4	NOS_{60}	1212.79±3.27	1023.41±4.58	828.69±4.78	928.11±5.17	729.89±2.97
5	NOS_{45}	1211.70±3.92	1021.50±4.26	826.47±7.43	918.30±6.78	717.88±4.55
6	NOA_{10}	1214.24±2.86	1022.13±4.48	827.22±5.00	921.39±5.32	732.23±7.00
7	AU_5	1209.98±4.72	1022.74±2.78	825.96±2.90	919.28±4.56	716.33±2.74
8	NOS_{60} + AU_5	1204.93±3.78	1017.53±2.28	826.32±1.79	917.41±6.01	713.89±4.32
9	NOA_{10} + AU_5	1205.99±3.81	1015.73±5.32	827.22±5.83	918.52±6.01	714.63±2.34
10	NOS_{60} + AU_5 + NOA_{10}	1200.89±3.10	1016.59±2.94	826.64±1.93	915.54±5.54	712.78±5.93
11	NOS_{45} + AU_5 + NOA_{10}	1203.06±7.00	1015.82±2.35	826.77±4.17	916.78±4.92	713.11±5.18

Three runs with different window sizes were performed in each sub-sampled data set and the mean and standard deviation of these runs are summarized.

D.4 User Response Statistics for Engagement and Challenge Questions

The table starts on next page.

User					Engagement								Challenge				
No	N	Pol.	Entr.	Acq.	Disacq.	MRS	PRS	NRS	Hist.	Pol.	Entr.	Acq.	Disacq.	MRS	PRS	NRS	Hist.
U01	31	.48	1.86	.24	.59	.18	.24	.44		.34	1.69	.15	.79	.06	.15	.47	
U02	15	.29	1.50	.18	.82	.00	.00	.41		.19	1.65	.06	.82	.12	.00	.35	
U03	45	.54	2.10	.54	.42	.04	.36	.24		.35	2.07	.26	.62	.12	.06	.20	
U04	31	.63	2.06	.55	.36	.09	.36	.30		.69	2.03	.52	.39	.09	.36	.33	
U05	24	.55	2.17	.46	.42	.13	.13	.29		.37	1.97	.33	.46	.21	.00	.25	
U06	37	.79	1.87	.46	.46	.08	.41	.38		.42	2.05	.27	.46	.27	.14	.41	
U07	32	.41	2.27	.31	.53	.16	.16	.28		.30	1.94	.19	.59	.22	.00	.38	
U08	13	.34	1.67	.69	.31	.00	.15	.08		.29	1.92	.23	.38	.38	.00	.15	
U09	14	.49	1.96	.43	.57	.00	.21	.36		.27	1.79	.07	.50	.43	.00	.29	
U10	36	.36	2.13	.22	.64	.14	.14	.36		.17	1.62	.06	.83	.11	.03	.56	
U11	27	.35	1.97	.30	.41	.30	.00	.22		.34	2.04	.22	.37	.41	.04	.22	
U12	14	.39	2.07	.29	.50	.21	.14	.43		.36	1.79	.21	.50	.29	.00	.43	
U13	50	.56	2.29	.42	.40	.18	.26	.24		.40	2.18	.24	.52	.24	.12	.36	
U14	49	.51	2.15	.37	.53	.10	.22	.39		.47	2.17	.35	.51	.14	.20	.39	

N: Number of responses, Pol.: Polarization, Entr.: Entropy, Acq.: Acquiescence, Disacq.: Disacquiescence, MRS: Middle Response Style
PRS: Positive Extreme Response Style, NRS: Negative Extreme Response Style, Hist.: Histogram

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Çavdar, Şeyma
Nationality: Turkish (TC)
Date and Place of Birth: 17.11.1989, Nazilli/AYDIN
Marital Status: Married

EDUCATION

Degree	Institution	Year of Graduation
Ph.D.	Information Systems, METU	2019
M.Sc.	Information Systems, METU	2013
B.Sc.	Industrial Engineering, TOBB UET	2011
High School	Nazilli Teacher Training High School	2007

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2011-2019	Informatics Institute, METU, Turkey	Research Assistant

PUBLICATIONS

1. **S. Kucukozer-Cavdar**, T. Taskaya-Temizel, A. Mehrotra, M. Musolesi, P. Tino. A Hybrid Model to Predict Office Workers' Availability for Rest Breaks Using Mobile Sensors, *Pervasive and Mobile Computing*, 2019. (Submitted)

2. **S. Kucukozer-Cavdar**, T. Taskaya-Temizel, M. Musolesi, P. Tino. A Multi-perspective Analysis of Social Context and Personal Factors in Office Settings for the Design of an Effective Mobile Notification System, *Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, 2019. (Submitted)

3. P.Unal, **S. Kucukozer-Cavdar**, T. Taskaya-Temizel, P. E. Eren, M. S. Iyengar. User Adoption and Evaluation of Mobile Health Applications: The Case for Physical Activity Monitoring, *Current and Emerging mHealth Technologies: Adoption, Implementation, and Use*, 2018.

4. P. Unal, **S. Kucukozer Cavdar**, T. Taskaya-Temizel, P. Erhan Eren, S.Iyengar. Exploring Behavior Change Features for Mobile Workout Applications, *MobiWIS*, 2017, Prague, Czech Republic.

5. S. Peker, **S. Kucukozer-Cavdar**, K. Cagiltay. Exploring the relationship between web presence and web usability for universities: A case study from Turkey, *Program*, 50(2), 2016, pp.157-174.

6. **S.Kucukozer-Cavdar**, T. Taskaya-Temizel. Analyzing the effects of the personality traits on the success of online collaborative groups, *Procedia-Social and Behavioral Sciences*, 228, 2016, pp.383-389.

7. **S. Kucukozer-Cavdar**, T. Taskaya-Temizel. A framework for investigating the factors that affect knowledge workers to take rest breaks in work hours, *8th International Conference on Occupational Safety&Health*, 8-11 May 2016, Istanbul, Turkey.

8. **S. Kucukozer-Cavdar**, E. Akkurt, D. Cavdar, A. Temizel, T. Taskaya-Temizel. UBDroid: Kullanıcı Davranış Analizi için Akıllı Telefon Uygulamaları Kullanım İzleme Aracı, *10. Ulusal Yazılım Mühendisliği Sempozyumu (UYMS)*, 24-26 October 2016, Canakkale, Turkey.